Adaptive inexact Newton methods for discretizations of nonlinear diffusion PDEs. II. Applications

Alexandre Ern, Martin Vohralík

Adaptive inexact Newton methods for discretizations of nonlinear diffusion PDEs. II. Applications*

Alexandre Ern* Martin Vohralík†

March 21, 2012

Abstract

We consider nonlinear algebraic systems resulting from numerical discretizations of nonlinear partial differential equations of diffusion type. In order to solve them, some iterative nonlinear solver, and, on each step of this solver, some iterative linear solver are used. In Part I of this work, we have developed a general abstract framework hinging on equilibrated flux reconstructions to derive stopping criteria for both iterative solvers and to control the size and distribution of the overall approximation error. In this Part II, we apply this framework to various discretization schemes like finite elements, nonconforming finite elements, discontinuous Galerkin, finite volumes, and lowest-order mixed finite elements; to different linearizations like fixed point and Newton; and to arbitrary iterative linear solvers. This leads to new guaranteed and robust a posteriori error estimates for nonlinear diffusion problems in the presence of linearization and algebraic errors. Moreover, for many discretization schemes, we improve on, or derive new, flux equilibration techniques.

Key words: nonlinear diffusion PDE, nonlinear algebraic system, adaptive linearization, adaptive algebraic solution, a posteriori error estimate, finite elements, nonconforming finite elements, discontinuous Galerkin, finite volumes, mixed finite elements

1 Introduction

In Part I of this work, we considered the following nonlinear diffusion problem: find a scalar-valued function u, termed the *potential*, such that

$$-\nabla\cdot\boldsymbol{\sigma}(\mathbf{x},u(\mathbf{x}),\nabla u(\mathbf{x})) = f \qquad \text{in } \Omega, \tag{1.1a}$$

$$u = 0 \qquad \text{on } \partial\Omega, \tag{1.1b}$$

where $\Omega \subset \mathbb{R}^d$, $d \geq 2$, is a polygonal (polyhedral) domain, $\boldsymbol{\sigma} : \Omega \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$, and $f : \Omega \to \mathbb{R}$ is the source term. Given a potential u, the vector-valued function $-\boldsymbol{\sigma}(\cdot, u, \nabla u) : \Omega \to \mathbb{R}^d$ is termed the *flux*. To simplify, we omit the dependence on the space variable $\mathbf{x}$ and simply write $\boldsymbol{\sigma}(u, \nabla u)$. Assuming that there is a real number $p \in (1, \infty)$ such that $f \in L^q(\Omega)$ with $q := \frac{p}{p-1}$, the model problem (1.1) is formulated as follows: find $u \in V := W_0^{1,p}(\Omega)$ such that

$$(\boldsymbol{\sigma}(u, \nabla u), \nabla v) = (f, v) \qquad \forall v \in V, \tag{1.2}$$

where, for $w \in L^q(\Omega)$, $v \in L^p(\Omega)$, (w, v) stands for $\int_\Omega w(\mathbf{x})v(\mathbf{x})\,\mathrm{d}\mathbf{x}$. In what follows, we assume that there exists a unique weak solution of (1.2).

*This work was partly supported by the Groupement MoMaS (PACEN/CNRS, ANDRA, BRGM, CEA, EdF, IRSN) and by the ERT project "Enhanced oil recovery and geological sequestration of CO_2: mesh adaptivity, a posteriori error control, and other advanced techniques" (LJLL/IFPEN).

*Université Paris-Est, CERMICS, Ecole des Ponts ParisTech, 77455 Marne la Vallée cedex 2, France (ern@cermics.enpc.fr).

†UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, 75005, Paris, France & CNRS, UMR 7598, Laboratoire Jacques-Louis Lions, 75005, Paris, France (vohralik@ann.jussieu.fr).

Problem (1.2) covers two important examples. Firstly, the *quasi-linear diffusion* problem where $\boldsymbol{\sigma}(v,\boldsymbol{\xi}) = \underline{\mathbf{A}}(v)\boldsymbol{\xi}$ with the assumption that the (tensor-valued) function $\underline{\mathbf{A}}$ is bounded and takes symmetric values with minimal eigenvalue uniformly bounded away from zero. In this case, $\boldsymbol{\sigma}$ depends linearly on $\boldsymbol{\xi}$, and the functional setting corresponds to the choice $p = 2$. Secondly, the *Leray–Lions* problem where $\boldsymbol{\sigma}$ depends nonlinearly on $\boldsymbol{\xi}$, but is independent of v, in the form $\boldsymbol{\sigma}(\boldsymbol{\xi}) = \underline{\mathbf{A}}(\boldsymbol{\xi})\boldsymbol{\xi}$ with suitable assumptions on the (tensor-valued) function $\underline{\mathbf{A}}$, see Part I. A typical example is the p-Laplacian where $\underline{\mathbf{A}}(\boldsymbol{\xi}) = |\boldsymbol{\xi}|^{p-2}\underline{\mathbf{I}}$ and $\underline{\mathbf{I}}$ is the identity tensor.

A posteriori estimates of discretization errors for problems of type (1.2) have been derived in various specific situations. Verfürth [36] developed a general framework for reliable and efficient a posteriori estimates in the finite element setting. Pousin and Rappaz [33] considered such ideas for general Petrov–Galerkin approximations. Repin [35] derived guaranteed a posteriori error estimates for arbitrary conforming discretizations; approximate solution of a global problem is, however, often necessary. Quite tight guaranteed upper bounds have been obtained by Carstensen and Klose [10] for the p-Laplacian. Other discretization schemes were also studied. Creusé *et al.* [13] derived a posteriori error estimates for mixed finite elements in the p-Laplacian case, whereas Houston *et al.* [26] considered the discontinuous Galerkin method in the quasi-linear diffusion setting. Kim [28] derived guaranteed estimates in the quasi-linear diffusion setting for locally conservative methods.

Recently, there has been an interest in setting up unified frameworks for a posteriori analysis. We mention, in particular, the work of Carstensen [9], Kim [28], and Ainsworth [1]. The last two references fall into the *flux equilibration* approach, which can be traced back to Prager and Synge [34], cf. also Luce and Wohlmuth [30], Braess and Schöberl [5], and the references therein. In [21], a unified analysis framework is developed whose application to a given discretization consists in the definition of flux reconstructions and the verification of a couple of assumptions on these fluxes.

The discretization of problem (1.2) leads to a system of nonlinear algebraic equations. In practice, this system is solved iteratively using a nonlinear solver which, at each step, employs an iterative linear solver. It is therefore important to distinguish and estimate separately the three *error components*, namely discretization, linearization, and algebraic errors. We have achieved this goal in Part I of this work, where we have proposed a posteriori error estimates for these components and balanced them through *stopping criteria* for the iterative nonlinear and linear solvers, leading to an *adaptive inexact Newton method*. Moreover, the obtained error estimates are *guaranteed* and *robust* with respect to the size of the nonlinearity owing to the chosen error measure, see (2.1) below. These developments are presented in a *unified abstract framework*. They extend the ideas of Chaillou and Suri [11, 12] and [18] concerning linearization errors, and those of [27] concerning algebraic errors. To our knowledge, this is the first time that the three error components are analyzed simultaneously. Alternative approaches include that of Han [24] for linearization errors and that of Becker *et al.* [4] and Arioli *et al.* [3] for algebraic errors, see also the references therein.

The aim of this Part II is to apply the framework of Part I to a wide class of discretization schemes, namely nonconforming finite elements, conforming finite elements, discontinuous Galerkin, finite volumes, and lowest-order mixed finite elements. In Section 2, we first synthesize the unified framework of Part I. The presentation somewhat differs from Part I so as to emphasize the key algorithmic aspects of the adaptive inexact Newton method and to identify the ingredients needed for its implementation together with the two key assumptions to be verified by the flux reconstructions. This verification is undertaken in Section 3–Section 7 for the various discretization schemes. For each scheme, we exemplify two iterative nonlinear solvers, namely fixed point and Newton, while the iterative linear solver can be arbitrary. It turns out that in many cases, we improve on, or develop new, flux equilibration techniques that are of independent interest. A brief discussion is included in the section dedicated to each discretization scheme. Finally, we draw some conclusions in Section 8.

2 The adaptive inexact Newton method

This section presents the basic setting and synthesizes the main results derived in Part I.

2.1 Basic notation

Let $\mathcal{T}_h$ be a simplicial mesh of Ω. For simplicity, we suppose that $\mathcal{T}_h$ does not contain hanging nodes, so that, for two distinct elements of $\mathcal{T}_h$, their intersection is either an empty set or a common l-dimensional face, $0 \le l \le d-1$. The $(d-1)$-dimensional faces of the mesh are collected in the set $\mathcal{E}_h$ such that $\mathcal{E}_h = \mathcal{E}_h^{\rm int} \cup \mathcal{E}_h^{\rm ext}$, with $\mathcal{E}_h^{\rm int}$ collecting interfaces and $\mathcal{E}_h^{\rm ext}$ boundary faces. The faces of a generic element $K \in \mathcal{T}_h$ are collected in the set $\mathcal{E}_K$. For any $K \in \mathcal{T}_h$ (resp., any $e \in \mathcal{E}_h$), h_K (resp., h_e) denotes its diameter. For any $K \in \mathcal{T}_h$, $\mathfrak{T}_K$ collects the elements $K' \in \mathcal{T}_h$ which share at least a vertex with K. Similarly, $\mathfrak{E}_K$ collects the faces which share at least a vertex with K, and we set $\mathfrak{E}_K^{\rm int} := \mathfrak{E}_K \cap \mathcal{E}_h^{\rm int}$. For any $e \in \mathcal{E}_h$, $\mathbf{n}_e$ stands for the unit normal vector to e (the orientation is irrelevant, but fixed, for all $e \in \mathcal{E}_h^{\rm int}$ and points outward Ω for all $e \in \mathcal{E}_h^{\rm ext}$) and, for any $K \in \mathcal{T}_h$, $\mathbf{n}_K$ stands for the outward unit normal vector to K.

We work with approximations that are possibly nonconforming, that is, not included in the space V. For this reason, we introduce the broken Sobolev space $V(\mathcal{T}_h) := \{v \in L^p(\Omega),\, v|_K \in W^{1,p}(K) \quad \forall K \in \mathcal{T}_h\}$. For any function $v \in V(\mathcal{T}_h)$, ∇v denotes its broken gradient, that is, the distributional gradient evaluated elementwise. Moreover, $[\![v]\!]$ denotes the difference (evaluated along $\mathbf{n}_e$) of the traces of v from the two adjacent mesh elements if $e \in \mathcal{E}_h^{\rm int}$ and the actual trace of v if $e \in \mathcal{E}_h^{\rm ext}$.

2.2 Approximate solution, approximate gradient, and error measure

Let a numerical scheme be used to discretize (1.2), and consider the k-th step, $k \ge 1$, of a nonlinear solver and the i-th step, $i \ge 1$, of a linear solver. We denote $u_h^{k,i}$ the corresponding discrete potential; we merely suppose that $u_h^{k,i} \in V(\mathcal{T}_h)$. Separately from $u_h^{k,i}$, we also consider a discrete gradient $\mathbf{g}_h^{k,i} \in [L^p(\Omega)]^d$. This allows us to handle a wide class of discretization schemes in a unified setting. For conforming schemes, $\mathbf{g}_h^{k,i}$ is obtained by applying the usual gradient to $u_h^{k,i}$; for various nonconforming schemes, the broken gradient can be used instead, but some schemes employ a more elaborate construction of $\mathbf{g}_h^{k,i}$, taking into account, e.g., the jumps of $u_h^{k,i}$. In all cases, whenever $u_h^{k,i} \in V$, there holds $\mathbf{g}_h^{k,i} = \nabla u_h^{k,i}$.

We measure the error as

$$\mathcal{J}_u(u_h^{k,i}, \mathbf{g}_h^{k,i}) = \mathcal{J}_{u,\rm F}(u_h^{k,i}, \mathbf{g}_h^{k,i}) + \mathcal{J}_{u,\rm NC}(u_h^{k,i}), \tag{2.1}$$

where

$$\mathcal{J}_{u,\rm F}(u_h^{k,i}, \mathbf{g}_h^{k,i}) := \sup_{\varphi \in V;\, \|\nabla\varphi\|_p = 1} \Big(\boldsymbol{\sigma}(u, \nabla u) - \boldsymbol{\sigma}(u_h^{k,i}, \mathbf{g}_h^{k,i}), \nabla\varphi \Big), \tag{2.2a}$$

$$\mathcal{J}_{u,\rm NC}(u_h^{k,i}) := \left\{ \sum_{K \in \mathcal{T}_h} \sum_{e \in \mathcal{E}_K} \alpha_e^s h_e^{1-s} \|[\![u - u_h^{k,i}]\!]\|_{s,e}^s \right\}^{\frac{1}{q}}. \tag{2.2b}$$

The quantity $\mathcal{J}_{u,\rm F}(u_h^{k,i}, \mathbf{g}_h^{k,i})$ measures the error in the fluxes (using a dual norm), see Chaillou and Suri [11, 12] and [18] for conforming discretizations. The quantity $\mathcal{J}_{u,\rm NC}(u_h^{k,i})$ measures the nonconformity of the discrete potential (recall that a function $v \in V(\mathcal{T}_h)$ is in V if and only if $[\![v]\!] = 0$ for all $e \in \mathcal{E}_h$, see, e.g., [17, Lemma 1.23]); a specific value for the weights α_e and the exponent $s \ge 1$ is only needed in Section 5.4. Owing to the well-posedness of (1.2), the characterization of conformity through jumps, and the above consistency of the discrete gradient in V, there holds $\mathcal{J}_u(u_h^{k,i}, \mathbf{g}_h^{k,i}) = 0$ if and only if $u = u_h^{k,i}$ and $\nabla u = \mathbf{g}_h^{k,i}$.

2.3 The algorithm

In Part I, we have derived an *adaptive inexact Newton method* to solve problems of the form

$$\mathcal{A}(U) = F, \tag{2.3}$$

where $\mathcal{A} : \mathbb{R}^N \to \mathbb{R}^N$ is a discrete nonlinear operator and $F \in \mathbb{R}^N$ a given vector, stemming from the discretization of problem (1.2) by a given numerical scheme. This algorithm is driven by stopping criteria based on a posteriori estimators distinguishing the three main error components, namely discretization, linearization, and algebraic errors. On a nonlinear solver step k, $k \ge 1$, and linear solver step i, $i \ge 1$, these estimators are respectively denoted by $\eta_{\rm disc}^{k,i}$, $\eta_{\rm lin}^{k,i}$, and $\eta_{\rm alg}^{k,i}$. A notion of algebraic remainder estimator

$\eta_{\mathrm{rem}}^{k,i}$ also appears. The evaluation of the estimators is discussed in Section 2.4. Let γ_{rem}, γ_{alg}, and γ_{lin} be positive user-given weights, typically of order 0.1. The algorithm reads:

Algorithm 2.1 (Adaptive inexact Newton method). *1. Choose an initial vector $U^0 \in \mathbb{R}^N$. Set $k := 1$.*

2. From U^{k-1}, define a matrix $\mathbb{A}^k \in \mathbb{R}^{N,N}$ and a vector $F^k \in \mathbb{R}^N$. Consider the following system of linear algebraic equations:

$$\mathbb{A}^k U^k = F^k. \tag{2.4}$$

3. (a) Define $U^{k,0} := U^{k-1}$ and set $i := 1$.

(b) Perform a step of a chosen iterative linear solver for the solution of the linear system (2.4), starting from the vector $U^{k,i-1}$. This yields an approximation $U^{k,i}$ to U^k which satisfies

$$\mathbb{A}^k U^{k,i} = F^k - R^{k,i}, \tag{2.5}$$

where $R^{k,i} \in \mathbb{R}^N$ is the algebraic residual vector on step i.

(c) Perform $\nu > 0$ additional steps of the iterative linear solver yielding an approximation $U^{k,i+\nu}$ to U^k which satisfies

$$\mathbb{A}^k U^{k,i+\nu} = F^k - R^{k,i+\nu}, \tag{2.6}$$

where $R^{k,i+\nu} \in \mathbb{R}^N$ is the algebraic residual vector on step $i+\nu$. The parameter ν is progressively increased until

$$\eta_{\mathrm{rem}}^{k,i} \le \gamma_{\mathrm{rem}} \max\{\eta_{\mathrm{disc}}^{k,i}, \eta_{\mathrm{lin}}^{k,i}, \eta_{\mathrm{alg}}^{k,i}\}. \tag{2.7}$$

(d) Check the convergence criterion for the linear solver in the form

$$\eta_{\mathrm{alg}}^{k,i} \le \gamma_{\mathrm{alg}} \max\{\eta_{\mathrm{disc}}^{k,i}, \eta_{\mathrm{lin}}^{k,i}\}. \tag{2.8}$$

If satisfied, set $U^k := U^{k,i}$. If not, set $i := i + \nu$ and go back to step 3b.

4. Check the convergence criterion for the nonlinear solver in the form

$$\eta_{\mathrm{lin}}^{k,i} \le \gamma_{\mathrm{lin}} \eta_{\mathrm{disc}}^{k,i}. \tag{2.9}$$

If satisfied, finish. If not, set $k := k + 1$ and go back to step 2.

In addition to the estimators $\eta_{\mathrm{disc}}^{k,i}$, $\eta_{\mathrm{lin}}^{k,i}$, $\eta_{\mathrm{alg}}^{k,i}$, and $\eta_{\mathrm{rem}}^{k,i}$ of Algorithm 2.1, we also introduced in Part I the data oscillation estimator $\eta_{\mathrm{osc}}^{k,i}$ and the quadrature estimator $\eta_{\mathrm{quad}}^{k,i}$. The evaluation of all the estimators is summarized in Section 2.4. In Part I, cf. Theorem 3.6, we derived the following guaranteed upper bound distinguishing the different error components:

$$\mathcal{J}_u(u_h^{k,i}, \mathbf{g}_h^{k,i}) \le \eta_{\mathrm{disc}}^{k,i} + \eta_{\mathrm{lin}}^{k,i} + \eta_{\mathrm{alg}}^{k,i} + \eta_{\mathrm{rem}}^{k,i} + \eta_{\mathrm{quad}}^{k,i} + \eta_{\mathrm{osc}}^{k,i}, \tag{2.10}$$

see also Theorem 3.4 in Part I for a sharper bound without the distinction of error components. Moreover, under the criteria (2.7), (2.8), and (2.9) in Algorithm 2.1, global efficiency and robustness was derived in the form, see Theorem 5.4 of Part I,

$$\eta_{\mathrm{disc}}^{k,i} + \eta_{\mathrm{lin}}^{k,i} + \eta_{\mathrm{alg}}^{k,i} + \eta_{\mathrm{rem}}^{k,i} \lesssim \mathcal{J}_u(u_h^{k,i}, \mathbf{g}_h^{k,i}) + \eta_{\mathrm{quad}}^{k,i} + \eta_{\mathrm{osc}}^{k,i}, \tag{2.11}$$

where $A \lesssim B$ stands for the inequality $A \le CB$ with a generic constant C independent of the mesh sizes h_K and h_e, the domain Ω, the nonlinear flux function $\boldsymbol{\sigma}$, and the Lebesgue exponent p, but depending on the shape regularity of the mesh family $\{\mathcal{T}_h\}_h$ and on the polynomial degrees of the discretization and the various reconstructions.

Remark 2.2 (Local stopping criteria and local efficiency). *Local, elementwise, stopping criteria are to be considered in (2.7), (2.8), and (2.9) in conjunction with adaptive mesh refinement. They yield local efficiency in a slightly different error measure, see Theorem 5.3 in Part I.*

2.4 Evaluation of the estimators

The evaluation of our estimators hinges on a few ingredients. First, we use three (vector-valued, piecewise polynomial) flux reconstructions $\mathbf{d}_h^{k,i}$, $\mathbf{l}_h^{k,i}$, and $\mathbf{a}_h^{k,i}$, where $\mathbf{d}_h^{k,i}$ is meant to approximate the discrete flux $-\boldsymbol{\sigma}(u_h^{k,i},\mathbf{g}_h^{k,i})$, $\mathbf{l}_h^{k,i}$ represents the linearization error, and $\mathbf{a}_h^{k,i}$ the algebraic error. Moreover, we use a (piecewise polynomial) function f_h to approximate the datum f, a (piecewise polynomial) function $\rho_h^{k,i}$ to represent the algebraic remainder, and a (vector-valued, piecewise polynomial) function $\overline{\boldsymbol{\sigma}}_h^{k,i}$ to approximate $\boldsymbol{\sigma}(u_h^{k,i},\mathbf{g}_h^{k,i})$. With these ingredients, the estimators can be evaluated as follows: for all $K\in\mathcal{T}_h$,

$$\eta_{\mathrm{disc},K}^{k,i} := 2^{1/p}\big(\|\overline{\boldsymbol{\sigma}}_h^{k,i}+\mathbf{d}_h^{k,i}\|_{q,K}+\eta_{\mathrm{NC},K}^{k,i}\big), \tag{2.12a}$$

$$\eta_{\mathrm{NC},K}^{k,i} := \left\{\sum_{e\in\mathcal{E}_K}\alpha_e^s h_e^{1-s}\|[\![u_h^{k,i}]\!]\|_{s,e}^s\right\}^{\frac1q}, \tag{2.12b}$$

$$\eta_{\mathrm{lin},K}^{k,i} := \|\mathbf{l}_h^{k,i}\|_{q,K},\qquad \eta_{\mathrm{alg},K}^{k,i} := \|\mathbf{a}_h^{k,i}\|_{q,K},\qquad \eta_{\mathrm{rem},K}^{k,i} := h_\Omega\|\rho_h^{k,i}\|_{q,K}, \tag{2.12c}$$

$$\eta_{\mathrm{osc},K}^{k,i} := C_{\mathrm{P},p}h_K\|f-f_h\|_{q,K},\qquad \eta_{\mathrm{quad},K}^{k,i} := \|\boldsymbol{\sigma}(u_h^{k,i},\mathbf{g}_h^{k,i})-\overline{\boldsymbol{\sigma}}_h^{k,i}\|_{q,K}, \tag{2.12d}$$

where h_Ω is the diameter of Ω and $C_{\mathrm{P},p}=\pi^{-\frac2p}d^{\frac12-\frac1p}$ for $p\ge2$ and $C_{\mathrm{P},p}=p^{\frac1p}2^{\frac{(p-1)}{p}}$ otherwise. The global versions of the estimators are $\eta_{\cdot}^{k,i} := \Big\{\sum_{K\in\mathcal{T}_h}\big(\eta_{\cdot,K}^{k,i}\big)^q\Big\}^{1/q}$.

2.5 Construction of the ingredients

The construction of our ingredients (that is, $\mathbf{d}_h^{k,i}$, $\mathbf{l}_h^{k,i}$, $\mathbf{a}_h^{k,i}$, f_h, $\rho_h^{k,i}$, and $\overline{\boldsymbol{\sigma}}_h^{k,i}$) proceeds in three steps. In the first step, we construct the sum $(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i})$, the function f_h, and the preliminary algebraic remainder $r_h^{k,i}$. The remainder $r_h^{k,i}$ is a piecewise polynomial constructed from the residual vector $R_h^{k,i}$ in (2.5), whereas the function f_h has to verify $(f_h,1)_K=(f,1)_K$ for all $K\in\mathcal{T}_h$. We require the following quasi-equilibration property:

Assumption 2.3 (Quasi-equilibration for $(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i})$)**.** *The function $(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i})$ is in $\mathbf{H}^q(\mathrm{div},\Omega)$ and satisfies*

$$\nabla\cdot(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i}) = f_h - r_h^{k,i}. \tag{2.13}$$

Our second step consists in considering the ν additional steps of the iterative linear solver (2.6). We then construct the fluxes $(\mathbf{d}_h^{k,i+\nu}+\mathbf{l}_h^{k,i+\nu})$ and remainder $r_h^{k,i+\nu}$ as in Assumption 2.3. We finally define the flux $\mathbf{a}_h^{k,i}$ and the algebraic remainder $\rho_h^{k,i}$ as

$$\mathbf{a}_h^{k,i} := (\mathbf{d}_h^{k,i+\nu}+\mathbf{l}_h^{k,i+\nu})-(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i}), \tag{2.14a}$$

$$\rho_h^{k,i} := r_h^{k,i+\nu}. \tag{2.14b}$$

Finally, our third step consists in specifying the reconstruction $\mathbf{d}_h^{k,i}$, and hence by subtraction the reconstruction $\mathbf{l}_h^{k,i}$, and the approximation $\overline{\boldsymbol{\sigma}}_h^{k,i}$. We let

$$\eta_{\sharp,\mathfrak{T}_K}^{k,i} := \left\{\sum_{K'\in\mathfrak{T}_K}h_{K'}^q\|f_h+\nabla\cdot\overline{\boldsymbol{\sigma}}_h^{k,i}\|_{q,K'}^q+\sum_{e\in\mathfrak{E}_K^{\mathrm{int}}}h_e\|[\![\overline{\boldsymbol{\sigma}}_h^{k,i}\cdot\mathbf{n}_e]\!]\|_{q,e}^q\right\}^{\frac1q}. \tag{2.15}$$

Set the shorthand notation $\eta_{\cdot,\mathfrak{T}_K}^{k,i} := \Big\{\sum_{K'\in\mathfrak{T}_K}\big(\eta_{\cdot,K'}^{k,i}\big)^q\Big\}^{\frac1q}$. The goal is to achieve:

Assumption 2.4 (Local approximation for $\mathbf{d}_h^{k,i}$ and $\overline{\boldsymbol{\sigma}}_h^{k,i}$ and convergence for $\mathbf{l}_h^{k,i}$)**.** *For all $K\in\mathcal{T}_h$, $\mathbf{d}_h^{k,i}$ and $\overline{\boldsymbol{\sigma}}_h^{k,i}$ satisfy the following local approximation property:*

$$\|\overline{\boldsymbol{\sigma}}_h^{k,i}+\mathbf{d}_h^{k,i}\|_{q,K}\lesssim\eta_{\sharp,\mathfrak{T}_K}^{k,i}+\eta_{\mathrm{NC},\mathfrak{T}_K}^{k,i}+\eta_{\mathrm{osc},\mathfrak{T}_K}^{k,i}. \tag{2.16}$$

Moreover, $\|\mathbf{l}_h^{k,i}\|_{q,K}\to0$ as the nonlinear solver converges.

Remark 2.5 (Tighter approximation property)**.** *In most cases, we actually prove* $\|\overline{\boldsymbol{\sigma}}_h^{k,i} + \mathbf{d}_h^{k,i}\|_{q,K} \lesssim \eta_{\sharp,\mathfrak{T}_K}^{k,i}$. $\eta_{\mathrm{osc},\mathfrak{T}_K}^{k,i}$ *appears for conforming finite elements in the lowest-order setting* $m = 1$ *and* $l = 0$*, see Section 4.4, while* $\eta_{\mathrm{NC},\mathfrak{T}_K}^{k,i}$ *appears for interior penalty discontinuous Galerkin when using full prescription for the flux reconstruction, see Section 5.4.*

In Section 3–Section 7, we show how to apply Algorithm 2.1 and the a posteriori error estimate (2.10) to various discretizations methods, to the fixed point and Newton linearizations, and to arbitrary iterative linear solvers. This consists in identifying the functional forms of (2.3), (2.4), and (2.5), the approximate solution $u_h^{k,i}$ and the approximate gradient $\mathbf{g}_h^{k,i}$, in constructing all the above ingredients, and in verifying Assumptions 2.3 and 2.4.

2.6 Flux reconstructions

We reconstruct the fluxes $\mathbf{d}_h^{k,i}$, $\mathbf{l}_h^{k,i}$, and $\mathbf{a}_h^{k,i}$ in the Raviart–Thomas–Nédélec spaces. We consider three approaches: (i) full prescription of the degrees of freedom, similarly to Destuynder and Métivet [14, 15]; (ii) solving local mixed finite element problems with prescribed boundary conditions, similarly to Luce and Wohlmuth [30] and to [20]; and (iii) solving local mixed finite element problems without prescribed boundary conditions, similarly to Braess and Schöberl [5]. The first approach is explicit, whereas the two other approaches are implicit, necessitating the solution of local linear systems, typically on a patch of elements.

For an integer $l \geq 0$, $\mathbb{P}_l(\mathcal{T}_h)$ denotes the broken polynomial space spanned by those functions v_h such that, for all $K \in \mathcal{T}_h$, $v_h|_K \in \mathbb{P}_l(K)$. For $\phi \in L^1(\Omega)$, $\Pi_l\phi \in \mathbb{P}_l(\mathcal{T}_h)$ is defined such that, for all $v_h \in \mathbb{P}_l(\mathcal{T}_h)$, $(\phi - \Pi_l\phi, v_h) = 0$; $\boldsymbol{\Pi}_l$ denotes the operator acting componentwise as Π_l on vector-valued functions. For $K \in \mathcal{T}_h$ and $l \geq 0$, let $\mathbf{RTN}_l(K) := [\mathbb{P}_l(K)]^d + \mathbf{x}\mathbb{P}_l(K)$ be the Raviart–Thomas–Nédélec finite element space of order l. Functions $\mathbf{v}_h \in \mathbf{RTN}_l(K)$ are such that, cf. Brezzi and Fortin [6], $\nabla{\cdot}\mathbf{v}_h \in \mathbb{P}_l(K)$ and $\mathbf{v}_h{\cdot}\mathbf{n}_e \in \mathbb{P}_l(e)$ for all $e \in \mathcal{E}_K$. We then set $\mathbf{RTN}_l^{-1}(\mathcal{T}_h) := \{\mathbf{v}_h \in [L^q(\Omega)]^d; \mathbf{v}_h|_K \in \mathbf{RTN}_l(K) \quad \forall K \in \mathcal{T}_h\}$ and $\mathbf{RTN}_l(\mathcal{T}_h) := \mathbf{RTN}_l^{-1}(\mathcal{T}_h) \cap \mathbf{H}^q(\mathrm{div}, \Omega)$. Functions in $\mathbf{RTN}_l(\mathcal{T}_h)$ have, in particular, a continuous normal component across interfaces. We use a similar notation for these spaces on various patches of elements. Finally, let $\mathbf{I}_l^{\mathrm{RTN}}$ stand for the broken Raviart–Thomas–Nédélec interpolation operator; for a smooth enough function $\mathbf{v}$, $\mathbf{I}_l^{\mathrm{RTN}}\mathbf{v} \in \mathbf{RTN}_l^{-1}(\mathcal{T}_h)$ is such that, for all $K \in \mathcal{T}_h$, letting $\langle w, v\rangle_e$ stand for $\int_e w(s)v(s)\,\mathrm{d}s$,

$$\langle(\mathbf{I}_l^{\mathrm{RTN}}\mathbf{v} - \mathbf{v})|_K{\cdot}\mathbf{n}_e, q_h\rangle_e = 0 \qquad \forall e \in \mathcal{E}_K,\ \forall q_h \in \mathbb{P}_l(e), \tag{2.17a}$$

$$(\mathbf{I}_l^{\mathrm{RTN}}\mathbf{v} - \mathbf{v}, \mathbf{r}_h)_K = 0 \qquad \forall \mathbf{r}_h \in [\mathbb{P}_{l-1}(K)]^d. \tag{2.17b}$$

3 Nonconforming finite elements

We treat here the discretization of problem (1.2) by nonconforming finite elements. We first present a simple elementwise flux reconstruction by full prescription. Then, we outline a slightly tighter reconstruction using a dual mesh, already considered in Section 6 of Part I, which is equivalent to solving local mixed finite element problems with prescription. Finally, we outline an equivalent viewpoint based on local mixed problems without prescription.

3.1 Discretization

Set $f_h := \Pi_0 f$. The nonconforming Crouzeix–Raviart finite element space V_h is spanned by piecewise affine polynomials on $\mathcal{T}_h$ such that the interface jumps and boundary face values have zero mean value over the corresponding face. The discretization of problem (1.2) reads: find $u_h \in V_h$ such that

$$(\boldsymbol{\sigma}(u_h, \nabla u_h), \nabla v_h) = (f_h, v_h) \qquad \forall v_h \in V_h. \tag{3.1}$$

The basis functions in V_h are associated with the interfaces and are denoted $\{\psi_e\}_{e\in\mathcal{E}_h^{\mathrm{int}}}$. Testing (3.1) against these functions yields the nonlinear algebraic system (2.3).

3.2 Linearization

Let $u_h^0 \in V_h$, fixing the initial vector U^0 in Algorithm 2.1. The linearization of (3.1), for $k \geq 1$, reads: find $u_h^k \in V_h$ such that

$$(\boldsymbol{\sigma}^{k-1}(u_h^k, \nabla u_h^k), \nabla \psi_e) = (f_h, \psi_e) \qquad \forall e \in \mathcal{E}_h^{\mathrm{int}}, \tag{3.2}$$

which is the functional form of the algebraic system (2.4). Two common ways to define the flux function $\boldsymbol{\sigma}^{k-1}$ are the fixed point linearization where

$$\boldsymbol{\sigma}^{k-1}(v, \boldsymbol{\xi}) := \underline{\mathbf{A}}(u_h^{k-1}, \nabla u_h^{k-1})\boldsymbol{\xi}, \tag{3.3}$$

and the Newton linearization where

$$\begin{aligned}\boldsymbol{\sigma}^{k-1}(v, \boldsymbol{\xi}) := \underline{\mathbf{A}}(u_h^{k-1}, \nabla u_h^{k-1})\boldsymbol{\xi} + (v - u_h^{k-1})\partial_v \underline{\mathbf{A}}(u_h^{k-1}, \nabla u_h^{k-1})\nabla u_h^{k-1} \\ + (\partial_{\boldsymbol{\xi}} \underline{\mathbf{A}}(u_h^{k-1}, \nabla u_h^{k-1}) \cdot \nabla u_h^{k-1}) \cdot (\boldsymbol{\xi} - \nabla u_h^{k-1}).\end{aligned} \tag{3.4}$$

3.3 Algebraic solution

On the i-th step, $i \geq 1$, of an iterative linear solver for the algebraic system (2.4), we obtain the algebraic residual vector $R^{k,i}$ in (2.5) with components associated with interfaces, $R^{k,i} = \{R_e^{k,i}\}_{e \in \mathcal{E}_h^{\mathrm{int}}}$. For convenience, we set $R_e^{k,i} := 0$ for all $e \in \mathcal{E}_h^{\mathrm{ext}}$. The functional form of (2.5) is: find $u_h^{k,i} \in V_h$ such that

$$(\boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i}), \nabla \psi_e) = (f_h, \psi_e) - R_e^{k,i} \qquad \forall e \in \mathcal{E}_h^{\mathrm{int}}. \tag{3.5}$$

3.4 Flux reconstruction by full prescription

Let $K \in \mathcal{T}_h$. We define $\mathbf{f}_h(\mathbf{x})|_K := \frac{f_h|_K}{d}(\mathbf{x} - \mathbf{x}_K)$, with $\mathbf{x}_K$ the barycenter of K. For all $e \in \mathcal{E}_K$, let $\mathbf{a}_{K,e}$ be the vertex of K opposite to the face e. Let $\mathcal{T}_e$ stand for the patch of elements sharing the face e. We first prescribe $(\mathbf{d}_h^{k,i} + \mathbf{l}_h^{k,i})$ elementwise in $\mathbf{RTN}_0^{-1}(\mathcal{T}_h)$ (as shown in Lemma 3.4 below, $(\mathbf{d}_h^{k,i} + \mathbf{l}_h^{k,i})$ turns out to be in $\mathbf{RTN}_0(\mathcal{T}_h)$).

Definition 3.1 (Construction of $(\mathbf{d}_h^{k,i} + \mathbf{l}_h^{k,i})$)**.** *Set, for all $K \in \mathcal{T}_h$,*

$$(\mathbf{d}_h^{k,i} + \mathbf{l}_h^{k,i})|_K := \big(-\boldsymbol{\Pi}_0 \boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i}) + \mathbf{f}_h\big)|_K - \sum_{e \in \mathcal{E}_K} |\mathcal{T}_e|^{-1} \frac{R_e^{k,i}}{d}(\mathbf{x} - \mathbf{a}_{K,e}). \tag{3.6}$$

The construction of $\mathbf{d}_h^{k,i}$ mimics that of $(\mathbf{d}_h^{k,i} + \mathbf{l}_h^{k,i})$ with $\boldsymbol{\sigma}(u_h^{k,i}, \nabla u_h^{k,i})$ in place of $\boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i})$. Specifically, let

$$\bar{R}_e^{k,i} := (f_h, \psi_e) - (\boldsymbol{\sigma}(u_h^{k,i}, \nabla u_h^{k,i}), \nabla \psi_e) \qquad \forall e \in \mathcal{E}_h^{\mathrm{int}}, \tag{3.7}$$

and $\bar{R}_e^{k,i} := 0$ for all $e \in \mathcal{E}_h^{\mathrm{ext}}$. We prescribe $\mathbf{d}_h^{k,i}$ (and hence, also $\mathbf{l}_h^{k,i}$ by subtraction):

Definition 3.2 (Construction of $\mathbf{d}_h^{k,i}$)**.** *Set, for all $K \in \mathcal{T}_h$,*

$$\mathbf{d}_h^{k,i}|_K := \big(-\boldsymbol{\Pi}_0 \boldsymbol{\sigma}(u_h^{k,i}, \nabla u_h^{k,i}) + \mathbf{f}_h\big)|_K - \sum_{e \in \mathcal{E}_K} |\mathcal{T}_e|^{-1} \frac{\bar{R}_e^{k,i}}{d}(\mathbf{x} - \mathbf{a}_{K,e}). \tag{3.8}$$

Definition 3.3 (Error measure, data oscillation, quadrature, and algebraic remainder)**.** *Use $u_h^{k,i}$ and $\mathbf{g}_h^{k,i} := \nabla u_h^{k,i}$ in the error measure (2.1) and set $f_h := \Pi_0 f$, $\overline{\boldsymbol{\sigma}}_h^{k,i} := \boldsymbol{\Pi}_0 \boldsymbol{\sigma}(u_h^{k,i}, \nabla u_h^{k,i})$, and $r_h^{k,i}|_K := \sum_{e \in \mathcal{E}_K} |\mathcal{T}_e|^{-1} R_e^{k,i}$ for all $K \in \mathcal{T}_h$.*

We now verify the assumptions of Section 2.5:

Lemma 3.4 (Quasi-equilibration)**.** *Assumption 2.3 holds.*

Proof. The proof exploits the link between nonconforming finite elements and mixed finite elements, cf. Marini [31] or Destuynder and Métivet [14]. For all $K \in \mathcal{T}_h$ and all $e \in \mathcal{E}_K$, we introduce the geometric weight $\omega_{e,K} := |K|/|\mathcal{T}_e|$. Note that $0 < \omega_{e,K} \le 1$ and $\omega_{e,K} = 1$ only on boundary faces. For any interface $e \in \mathcal{E}_h^{\rm int}$ such that $e = \partial K \cap \partial K'$, $K, K' \in \mathcal{T}_h$, observing that $\omega_{e,K} + \omega_{e,K'} = 1$, we define the weighted average of a piecewise polynomial function $\mathbf{v}_h$ at e as $\{\!\{\mathbf{v}_h\}\!\}_\omega := \omega_{e,K'}(\mathbf{v}_h|_K)|_e + \omega_{e,K}(\mathbf{v}_h|_{K'})|_e$. On boundary faces $e \in \mathcal{E}_h^{\rm ext}$, we set $\{\!\{\mathbf{v}_h\}\!\}_\omega := \mathbf{v}_h|_e$. We first show that, for all for all $K \in \mathcal{T}_h$ and all $e \in \mathcal{E}_K$,

$$(\mathbf{d}_h^{k,i} + \mathbf{l}_h^{k,i})|_K \cdot \mathbf{n}_e = \{\!\{-\boldsymbol{\Pi}_0 \boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i}) + \mathbf{f}_h\}\!\}_\omega \cdot \mathbf{n}_e. \tag{3.9}$$

This is obvious for $e \in \mathcal{E}_h^{\rm ext}$. Let now $e \in \mathcal{E}_h^{\rm int}$. Set $\mathbf{w}_h := -\boldsymbol{\Pi}_0 \boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i}) + \mathbf{f}_h$. It is readily seen that $(\boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i}), \nabla \psi_e) = |e| [\![\boldsymbol{\Pi}_0 \boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i})]\!] \cdot \mathbf{n}_e$ and $(f_h, \psi_e) = |e| [\![\mathbf{f}_h]\!] \cdot \mathbf{n}_e$ (recall that $[\![\cdot]\!]$ denotes the jump across e in the direction of $\mathbf{n}_e$). Hence, owing to (3.5), $[\![\mathbf{w}_h]\!] \cdot \mathbf{n}_e = |e|^{-1} R_e^{k,i}$. The result (3.9) then follows from

$$\mathbf{w}_h|_K \cdot \mathbf{n}_e = \{\!\{\mathbf{w}_h\}\!\}_\omega \cdot \mathbf{n}_e + \omega_{e,K} [\![\mathbf{w}_h]\!] \cdot \mathbf{n}_K \tag{3.10}$$

and (3.6). Now, (3.9) shows that $(\mathbf{d}_h^{k,i} + \mathbf{l}_h^{k,i})$ has continuous normal component across interfaces, so that $(\mathbf{d}_h^{k,i} + \mathbf{l}_h^{k,i}) \in \mathbf{RTN}_0(\mathcal{T}_h)$. Finally, the property (2.13) follows by taking the divergence of (3.6) and considering the definition of $r_h^{k,i}$. □

Lemma 3.5 (Local approximation and convergence)**.** *Assumption 2.4 holds.*

Proof. The requirement on $\mathbf{l}_h^{k,i}$ is obvious from Definitions 3.1 and 3.2. Turning to $\mathbf{d}_h^{k,i}$, we let $\mathbf{v}_h := \overline{\boldsymbol{\sigma}}_h^{k,i} + \mathbf{d}_h^{k,i} \in \mathbf{RTN}_0^{-1}(\mathcal{T}_h)$ and use, for all $K \in \mathcal{T}_h$, the estimate $\|\mathbf{v}_h\|_{q,K} \lesssim \{\sum_{e \in \mathcal{E}_K} h_e \|\mathbf{v}_h|_K \cdot \mathbf{n}_e\|_{q,e}^q\}^{\frac{1}{q}}$ shown in [18, Section A.4]. Let $e \in \mathcal{E}_K$. If $e \in \mathcal{E}_h^{\rm ext}$, using $\bar{R}_e^{k,i} := 0$ in (3.8), $|\mathbf{x} - \mathbf{x}_K| \le h_K$, a q-robust inverse inequality (see [18, Section A.1 and A.4]), the fact that f_h is constant on K, and $\nabla \cdot \overline{\boldsymbol{\sigma}}_h^{k,i} = 0$ yields

$$\begin{aligned} h_e \|\mathbf{v}_h|_K \cdot \mathbf{n}_e\|_{q,e}^q &= h_e \|f_h|_K d^{-1}(\mathbf{x} - \mathbf{x}_K) \cdot \mathbf{n}_e\|_{q,e}^q \le h_K^{1+q} \|f_h|_K\|_{q,e}^q \\ &\lesssim h_K^q \|f_h\|_{q,K}^q = h_K^q \|f_h + \nabla \cdot \overline{\boldsymbol{\sigma}}_h^{k,i}\|_{q,K}^q. \end{aligned}$$

If $e \in \mathcal{E}_h^{\rm int}$, reasoning as in the proof of Lemma 3.4 yields $\mathbf{d}_h^{k,i} \cdot \mathbf{n}_e = \{\!\{-\overline{\boldsymbol{\sigma}}_h^{k,i} + \mathbf{f}_h\}\!\}_\omega \cdot \mathbf{n}_e$ (so that $\mathbf{d}_h^{k,i} \in \mathbf{RTN}_0(\mathcal{T}_h)$). Using this relation, (3.10) to evaluate $\mathbf{v}_h|_K \cdot \mathbf{n}_e$, and the continuity of the normal component of $\mathbf{d}_h^{k,i}$ yields $\mathbf{v}_h|_K \cdot \mathbf{n}_e = \{\!\{\mathbf{f}_h\}\!\}_\omega \cdot \mathbf{n}_e + \omega_{e,K} [\![\overline{\boldsymbol{\sigma}}_h^{k,i}]\!] \cdot \mathbf{n}_K$. We conclude by proceeding as in the first part of the proof. □

3.5 Flux reconstruction by local mixed problems with prescription

A slightly tighter flux reconstruction was considered in Section 6 of Part I. For all $K \in \mathcal{T}_h$ and all $e \in \mathcal{E}_K$, let K_e be the sub-simplex of K formed by the face e and the barycenter $\mathbf{x}_K$. Let D_e regroup the two (or one for boundary faces) sub-simplices which share e. Then, the tighter flux reconstruction consists in replacing in the last terms of (3.6) and (3.8) the vertex $\mathbf{a}_{K,e}$ by the barycenter $\mathbf{x}_K$ and $|\mathcal{T}_e|^{-1}$ by $|D_e|^{-1}$. The advantage is that, using local stopping criteria, elementwise efficiency (without neighbors) can be proven on each element of the dual mesh $\mathcal{D}_h = \{D_e\}_{e \in \mathcal{E}_h}$. Moreover, for all $D_e \in \mathcal{D}_h$ with outward normal $\mathbf{n}_{D_e}$, letting $\mathcal{S}_{D_e}$ collect the sub-simplices in D_e,

$$(\mathbf{d}_h^{k,i} + \mathbf{l}_h^{k,i})|_{D_e} = \arg \inf_{\{\mathbf{v}_h \in \mathbf{RTN}_0^{\rm N}(\mathcal{S}_{D_e}),\, \nabla \cdot \mathbf{v}_h = f_h - r_h^{k,i}\}} \|\boldsymbol{\Pi}_0 \boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i}) + \mathbf{v}_h\|_{D_e},$$

where $\mathbf{RTN}_0^{\rm N}(\mathcal{S}_{D_e})$ fixes to $-\big(\boldsymbol{\Pi}_0 \boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i})\big) \cdot \mathbf{n}_{D_e}$ the normal component on those faces of ∂D_e that are inside Ω, see [20], [21, Section 4.5], or [25, Section 7.3], and $r_h^{k,i}|_{D_e} = R_e^{k,i} |D_e|^{-1}$. This construction stems from the face-centered finite volume method, where the equivalent of (3.5), yielding the same solution $u_h^{k,i}$, is,

$$-\langle \boldsymbol{\Pi}_0 \boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i}) \cdot \mathbf{n}_{D_e}, 1 \rangle_{\partial D_e} = (f_h, 1)_{D_e} - R_e^{k,i} \qquad \forall e \in \mathcal{E}_h^{\rm int}. \tag{3.11}$$

A similar construction can be devised for $\mathbf{d}_h^{k,i}$.

3.6 Flux reconstruction by local mixed problems without prescription

We finally present a construction inspired by the approach of Braess and Schöberl [5] for conforming finite elements, see Section 4.4. For all $e \in \mathcal{E}_h$, let $\mathbf{RTN}_0^{\mathrm{N},0}(\mathcal{T}_e)$ denote the subspace of $\mathbf{RTN}_0(\mathcal{T}_e)$ with zero normal flux through $\partial\mathcal{T}_e$ for $e \in \mathcal{E}_h^{\mathrm{int}}$ and through that part of the boundary of $\mathcal{T}_e$ which lies inside Ω for $e \in \mathcal{E}_h^{\mathrm{ext}}$. Let $\mathbb{P}_0^*(\mathcal{T}_e)$ be spanned by piecewise constants on $\mathcal{T}_e$ with zero mean on $\mathcal{T}_e$ when $e \in \mathcal{E}_h^{\mathrm{int}}$; when $e \in \mathcal{E}_h^{\mathrm{ext}}$, the mean value condition is not imposed. Define $(\mathbf{d}_e^{k,i} + \mathbf{l}_e^{k,i}) \in \mathbf{RTN}_0^{\mathrm{N},0}(\mathcal{T}_e)$ and $q_e \in \mathbb{P}_0^*(\mathcal{T}_e)$ by the solution of the following mixed finite element problem on $\mathcal{T}_e$:

$$
\begin{aligned}
&(\mathbf{d}_e^{k,i} + \mathbf{l}_e^{k,i}, \mathbf{v}_h)_{\mathcal{T}_e} - (q_e, \nabla\cdot\mathbf{v}_h)_{\mathcal{T}_e} = -(\psi_e \boldsymbol{\Pi}_0 \boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i}), \mathbf{v}_h)_{\mathcal{T}_e},\\
&(\nabla\cdot(\mathbf{d}_e^{k,i} + \mathbf{l}_e^{k,i}), \phi_h)_{\mathcal{T}_e} = (f_h\psi_e - \boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i})\cdot\nabla\psi_e, \phi_h)_{\mathcal{T}_e} - (R_e^{k,i}, \phi_h)_{\mathcal{T}_e}|\mathcal{T}_e|^{-1},
\end{aligned}
$$

for all $(\mathbf{v}_h, \phi_h) \in \mathbf{RTN}_0^{\mathrm{N},0}(\mathcal{T}_e) \times \mathbb{P}_0^*(\mathcal{T}_e)$. Then, set $\mathbf{d}_h^{k,i} + \mathbf{l}_h^{k,i} := \sum_{e\in\mathcal{E}_h}(\mathbf{d}_e^{k,i} + \mathbf{l}_e^{k,i})$. In the above problems, we can take $\phi_h \in \mathbb{P}_0(\mathcal{T}_e)$ since (3.5) yields, for all $e \in \mathcal{E}_h^{\mathrm{int}}$, the Neumann compatibility condition

$$
(f_h, \psi_e)_{\mathcal{T}_e} - (\boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i}), \nabla\psi_e)_{\mathcal{T}_e} - R_e^{k,i} = 0. \tag{3.12}
$$

It can be shown that the constructions of $(\mathbf{d}_h^{k,i} + \mathbf{l}_h^{k,i})$ from Definition 3.1 and the present one coincide. A similar construction can be devised for $\mathbf{d}_h^{k,i}$.

4 Conforming finite elements

We treat here the discretization of problem (1.2) by conforming finite elements. We focus on flux reconstruction through local mixed problems without prescription, following Braess and Schöberl [5], cf. also Destuynder and Métivet [15]. A slightly tighter construction without prescription is possible in the piecewise affine case. Similarly to Section 3.5, this construction exploits the links between the piecewise affine finite element method and the vertex-centered finite volume method and allows for a fully local statement (without neighbors) of the local efficiency result on a vertex-centered dual mesh, cf. Luce and Wohlmuth [30] and [18].

4.1 Discretization

Let $V_h := \mathbb{P}_m(\mathcal{T}_h) \cap V$, $m \geq 1$, be the usual finite element space of continuous, piecewise m-th order polynomial functions. The corresponding discretization of problem (1.2) reads: find $u_h \in V_h$ such that

$$
(\boldsymbol{\sigma}(u_h, \nabla u_h), \nabla v_h) = (f_h, v_h) \qquad \forall v_h \in V_h. \tag{4.1}
$$

Let $\psi_j \in V_h$, $j \in \mathcal{C} := \{1, \dots, \dim(V_h)\}$, denote the basis functions of V_h. Employing these functions in (4.1) gives rise to the nonlinear algebraic system (2.3).

4.2 Linearization

Let $u_h^0 \in V_h$, fixing the initial vector U^0 in Algorithm 2.1. The linearization of (4.1), for $k \geq 1$, reads: find $u_h^k \in V_h$ such that

$$
(\boldsymbol{\sigma}^{k-1}(u_h^k, \nabla u_h^k), \nabla\psi_j) = (f, \psi_j) \qquad \forall j \in \mathcal{C}, \tag{4.2}
$$

which is the functional form of the algebraic system (2.4). Two common ways to define the flux function $\boldsymbol{\sigma}^{k-1}$ are the fixed point linearization (3.3) and the Newton linearization (3.4).

4.3 Algebraic solution

On the i-th step, $i \geq 1$, of an iterative linear solver for the algebraic system (2.4), we obtain the algebraic residual vector $R^{k,i}$ in (2.5), with components associated with the set $\mathcal{C}$, $R^{k,i} = \{R_j^{k,i}\}_{j\in\mathcal{C}}$. The functional form of (2.5) is: find $u_h^{k,i} \in V_h$ such that

$$
(\boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i}), \nabla\psi_j) = (f, \psi_j) - R_j^{k,i} \qquad \forall j \in \mathcal{C}. \tag{4.3}
$$

4.4 Flux reconstruction by local mixed problems without prescription

We construct $(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i}) \in \mathbf{RTN}_l(\mathcal{T}_h)$ with $l := m-1$ or $l := m$. The construction uses the partition of unity together with local mixed finite element problems posed on patches around mesh vertices, as outlined in Section 3.6. Let $\mathcal{V}_h$ denote the set of mesh vertices with subsets $\mathcal{V}_h^{\rm int}$ for interior vertices and $\mathcal{V}_h^{\rm ext}$ for boundary ones. Let $\psi_{\mathbf{a}} \in \mathbb{P}_1(\mathcal{T}_h) \cap C^0(\Omega)$ stand for the classical hat basis function associated with vertex $\mathbf{a} \in \mathcal{V}_h$. To distribute the algebraic residual onto vertices, we set, for all $\mathbf{a} \in \mathcal{V}_h^{\rm int}$, $R_{\mathbf{a}}^{k,i} := \sum_{j\in\mathcal{C}} \beta_j R_j^{k,i}$ where the coefficients β_j are such that $\psi_{\mathbf{a}} = \sum_{j\in\mathcal{C}} \beta_j \psi_j$, while, for $\mathbf{a} \in \mathcal{V}_h^{\rm ext}$, we set $R_{\mathbf{a}}^{k,i} := 0$. Furthermore, for all $\mathbf{a} \in \mathcal{V}_h$, let $\mathcal{T}_{\mathbf{a}}$ be the patch of elements of $\mathcal{T}_h$ that share $\mathbf{a}$, and let $\mathbf{RTN}_l^{\rm N,0}(\mathcal{T}_{\mathbf{a}})$ be the subspace of $\mathbf{RTN}_l(\mathcal{T}_{\mathbf{a}})$ with zero normal flux through $\partial\mathcal{T}_{\mathbf{a}}$ for $\mathbf{a} \in \mathcal{V}_h^{\rm int}$ and through that part of $\partial\mathcal{T}_{\mathbf{a}}$ which lies inside Ω for $\mathbf{a} \in \mathcal{V}_h^{\rm ext}$. Let $\mathbb{P}_l^*(\mathcal{T}_{\mathbf{a}})$ be spanned by piecewise l-th order polynomials on $\mathcal{T}_{\mathbf{a}}$ with zero mean on $\mathcal{T}_{\mathbf{a}}$ when $\mathbf{a} \in \mathcal{V}_h^{\rm int}$; when $\mathbf{a} \in \mathcal{V}_h^{\rm ext}$, the mean value condition is not imposed.

Definition 4.1 (Construction of $(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i})$)**.** *For all vertices $\mathbf{a} \in \mathcal{V}_h$, define $(\mathbf{d}_{\mathbf{a}}^{k,i}+\mathbf{l}_{\mathbf{a}}^{k,i}) \in \mathbf{RTN}_l^{\rm N,0}(\mathcal{T}_{\mathbf{a}})$ and $q_{\mathbf{a}} \in \mathbb{P}_l^*(\mathcal{T}_{\mathbf{a}})$ by*

$$(\mathbf{d}_{\mathbf{a}}^{k,i}+\mathbf{l}_{\mathbf{a}}^{k,i},\mathbf{v}_h)_{\mathcal{T}_{\mathbf{a}}} - (q_{\mathbf{a}},\nabla\cdot\mathbf{v}_h)_{\mathcal{T}_{\mathbf{a}}} = -(\mathbf{I}_l^{\rm RTN}(\psi_{\mathbf{a}}\boldsymbol{\Pi}_l\boldsymbol{\sigma}^{k-1}(u_h^{k,i},\nabla u_h^{k,i})),\mathbf{v}_h)_{\mathcal{T}_{\mathbf{a}}}, \tag{4.4a}$$

$$(\nabla\cdot(\mathbf{d}_{\mathbf{a}}^{k,i}+\mathbf{l}_{\mathbf{a}}^{k,i}),\phi_h)_{\mathcal{T}_{\mathbf{a}}} = (f\psi_{\mathbf{a}} - \boldsymbol{\sigma}^{k-1}(u_h^{k,i},\nabla u_h^{k,i})\cdot\nabla\psi_{\mathbf{a}},\phi_h)_{\mathcal{T}_{\mathbf{a}}} - (R_{\mathbf{a}}^{k,i},\phi_h)_{\mathcal{T}_{\mathbf{a}}}|\mathcal{T}_{\mathbf{a}}|^{-1}, \tag{4.4b}$$

for all $(\mathbf{v}_h,\phi_h) \in \mathbf{RTN}_l^{\rm N,0}(\mathcal{T}_{\mathbf{a}}) \times \mathbb{P}_l^*(\mathcal{T}_{\mathbf{a}})$. Then, set $\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i} := \sum_{\mathbf{a}\in\mathcal{V}_h}(\mathbf{d}_{\mathbf{a}}^{k,i}+\mathbf{l}_{\mathbf{a}}^{k,i})$.

In the above problems, we can take $\phi_h \in \mathbb{P}_l(\mathcal{T}_{\mathbf{a}})$ since multiplying (4.3) by the coefficients β_j, summing over all $j \in \mathcal{C}$, and using the definition of $R_{\mathbf{a}}^{k,i}$, yields, for all $\mathbf{a} \in \mathcal{V}_h^{\rm int}$, the Neumann compatibility condition

$$(\boldsymbol{\sigma}^{k-1}(u_h^{k,i},\nabla u_h^{k,i}),\nabla\psi_{\mathbf{a}})_{\mathcal{T}_{\mathbf{a}}} = (f,\psi_{\mathbf{a}})_{\mathcal{T}_{\mathbf{a}}} - R_{\mathbf{a}}^{k,i}. \tag{4.5}$$

Moreover, it can be shown that the construction of Definition 4.1 is equivalent to the one proposed by Braess and Schöberl [5]. We proceed similarly for $\mathbf{d}_h^{k,i}$. Set

$$\bar{R}_{\mathbf{a}}^{k,i} := (f,\psi_{\mathbf{a}})_{\mathcal{T}_{\mathbf{a}}} - (\boldsymbol{\sigma}(u_h^{k,i},\nabla u_h^{k,i}),\nabla\psi_{\mathbf{a}})_{\mathcal{T}_{\mathbf{a}}} \qquad \forall \mathbf{a} \in \mathcal{V}_h^{\rm int}, \tag{4.6}$$

and for any $\mathbf{a} \in \mathcal{V}_h^{\rm ext}$, set $\bar{R}_{\mathbf{a}}^{k,i} := 0$.

Definition 4.2 (Construction of $\mathbf{d}_h^{k,i}$)**.** *Define $\mathbf{d}_{\mathbf{a}}^{k,i} \in \mathbf{RTN}_l^{\rm N,0}(\mathcal{T}_{\mathbf{a}})$ and $\bar{q}_{\mathbf{a}} \in \mathbb{P}_l^*(\mathcal{T}_{\mathbf{a}})$ by solving the mixed finite element problems (4.4) with $\boldsymbol{\sigma}(u_h^{k,i},\nabla u_h^{k,i})$ in place of $\boldsymbol{\sigma}^{k-1}(u_h^{k,i},\nabla u_h^{k,i})$ and $\bar{R}_{\mathbf{a}}^{k,i}$ in place of $R_{\mathbf{a}}^{k,i}$. Then, set $\mathbf{d}_h^{k,i} := \sum_{\mathbf{a}\in\mathcal{V}_h}\mathbf{d}_{\mathbf{a}}^{k,i}$.*

Definition 4.3 (Error measure, data oscillation, quadrature, and algebraic remainder)**.** *Use $u_h^{k,i}$ and $\mathbf{g}_h^{k,i} := \nabla u_h^{k,i}$ in the error measure (2.1) and set $f_h := \Pi_l f$, $\overline{\boldsymbol{\sigma}}_h^{k,i} := \boldsymbol{\Pi}_l\boldsymbol{\sigma}(u_h^{k,i},\nabla u_h^{k,i})$, and $r_h^{k,i}|_K := \sum_{\mathbf{a}\in\mathcal{V}_K}|\mathcal{T}_{\mathbf{a}}|^{-1}R_{\mathbf{a}}^{k,i}$ for all $K \in \mathcal{T}_h$, where $\mathcal{V}_K$ collects the vertices of the element K.*

We now verify the assumptions of Section 2.5:

Lemma 4.4 (Quasi-equilibration)**.** *Assumption 2.3 holds.*

Proof. Let $K \in \mathcal{T}_h$ and let $v_h \in \mathbb{P}_l(K)$ (and zero elsewhere) be fixed. For any $\mathbf{a} \in \mathcal{V}_K$, by (4.5), we can take v_h as test function ϕ_h in (4.4b). Since $\sum_{\mathbf{a}\in\mathcal{V}_K}\psi_{\mathbf{a}}|_K = 1$ and $\sum_{\mathbf{a}\in\mathcal{V}_K}\nabla\psi_{\mathbf{a}}|_K = 0$ ($\psi_{\mathbf{a}}$ form a partition of unity on K), we infer

$$(\nabla\cdot(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i}),v_h)_K = \sum_{\mathbf{a}\in\mathcal{V}_K}(\nabla\cdot(\mathbf{d}_{\mathbf{a}}^{k,i}+\mathbf{l}_{\mathbf{a}}^{k,i}),v_h)_K = (f,v_h)_K - \sum_{\mathbf{a}\in\mathcal{V}_K}(R_{\mathbf{a}}^{k,i},v_h)_K|\mathcal{T}_{\mathbf{a}}|^{-1},$$

whence the assertion of the lemma follows from the definition of $r_h^{k,i}$. □

Lemma 4.5 (Local approximation and convergence)**.** *Assumption 2.4 holds.*

Proof. The requirement on $\mathbf{l}_h^{k,i}$ is obvious from Definitions 4.1 and 4.2. Turning to $\mathbf{d}_h^{k,i}$, let $K \in \mathcal{T}_h$. Since $\mathbf{I}_l^{\mathrm{RTN}}(\overline{\boldsymbol{\sigma}}_h^{k,i}) = \overline{\boldsymbol{\sigma}}_h^{k,i}$, by the partition of unity and linearity of the projection operator $\mathbf{I}_l^{\mathrm{RTN}}$, it follows that $(\mathbf{d}_h^{k,i} + \overline{\boldsymbol{\sigma}}_h^{k,i})|_K = (\mathbf{d}_h^{k,i} + \mathbf{I}_l^{\mathrm{RTN}}(\overline{\boldsymbol{\sigma}}_h^{k,i}))|_K = \sum_{\mathbf{a}\in\mathcal{V}_K}(\mathbf{d}_{\mathbf{a}}^{k,i} + \mathbf{I}_l^{\mathrm{RTN}}(\psi_{\mathbf{a}}\overline{\boldsymbol{\sigma}}_h^{k,i}))|_K$. We thus only work with $(\mathbf{d}_{\mathbf{a}}^{k,i} + \mathbf{I}_l^{\mathrm{RTN}}(\psi_{\mathbf{a}}\overline{\boldsymbol{\sigma}}_h^{k,i}))|_K$ for one vertex $\mathbf{a} \in \mathcal{V}_K$, or, more precisely, with $(\mathbf{d}_{\mathbf{a}}^{k,i} + \mathbf{I}_l^{\mathrm{RTN}}(\psi_{\mathbf{a}}\overline{\boldsymbol{\sigma}}_h^{k,i}))|_{\mathcal{T}_{\mathbf{a}}}$, in order to prove (2.16). Note that $(\boldsymbol{\sigma}(u_h^{k,i}, \nabla u_h^{k,i}), \nabla\psi_{\mathbf{a}})_{\mathcal{T}_{\mathbf{a}}} = (\overline{\boldsymbol{\sigma}}_h^{k,i}, \nabla\psi_{\mathbf{a}})_{\mathcal{T}_{\mathbf{a}}}$ and, for all $\phi_h \in \mathbb{P}_l(\mathcal{T}_{\mathbf{a}})$, $(\boldsymbol{\sigma}(u_h^{k,i}, \nabla u_h^{k,i})\cdot\nabla\psi_{\mathbf{a}}, \phi_h)_{\mathcal{T}_{\mathbf{a}}} = (\overline{\boldsymbol{\sigma}}_h^{k,i}\cdot\nabla\psi_{\mathbf{a}}, \phi_h)_{\mathcal{T}_{\mathbf{a}}}$, so that we can replace $\boldsymbol{\sigma}(u_h^{k,i}, \nabla u_h^{k,i})$ by $\overline{\boldsymbol{\sigma}}_h^{k,i}$ everywhere in Definition 4.2. We next proceed as in [18, Section A.4], cf. also [25, Proof of Lemmas 7.5 and 7.8]. Firstly, let $M(\mathcal{T}_{\mathbf{a}})$ denote the postprocessing space of Arbogast and Chen [2]. This is a space of piecewise (discontinuous) polynomials m_h on $\mathcal{T}_{\mathbf{a}}$ such that

$$\langle [\![m_h]\!], v_h\rangle_e = 0 \qquad \forall e \in \mathcal{E}_{\mathbf{a}},\ \forall v_h \in \mathbb{P}_l(e), \tag{4.7}$$

where $\mathcal{E}_{\mathbf{a}}$ collects the faces to which $\mathbf{a}$ belongs. Moreover, the functions m_h in $M(\mathcal{T}_{\mathbf{a}})$ satisfy the condition $(m_h, 1)_{\mathcal{T}_{\mathbf{a}}} = 0$ if the vertex $\mathbf{a}$ lies in the interior of Ω. [39, Lemma 5.4] and the generalizations of [18, Section A.4] to the $L^q(\Omega)$-setting yield

$$\|\mathbf{d}_{\mathbf{a}}^{k,i} + \mathbf{I}_l^{\mathrm{RTN}}(\psi_{\mathbf{a}}\overline{\boldsymbol{\sigma}}_h^{k,i})\|_{q,\mathcal{T}_{\mathbf{a}}} \lesssim \sup_{m_h\in M(\mathcal{T}_{\mathbf{a}}),\, \|\nabla m_h\|_{p,\mathcal{T}_{\mathbf{a}}}=1} (\mathbf{d}_{\mathbf{a}}^{k,i} + \mathbf{I}_l^{\mathrm{RTN}}(\psi_{\mathbf{a}}\overline{\boldsymbol{\sigma}}_h^{k,i}), \nabla m_h)_{\mathcal{T}_{\mathbf{a}}}.$$

Let $m_h \in M(\mathcal{T}_{\mathbf{a}})$ with $\|\nabla m_h\|_{p,\mathcal{T}_{\mathbf{a}}} = 1$ be fixed and consider the right-hand side of the above inequality. The Green theorem, the fact that $\mathbf{d}_{\mathbf{a}}^{k,i} + \mathbf{I}_l^{\mathrm{RTN}}(\psi_{\mathbf{a}}\overline{\boldsymbol{\sigma}}_h^{k,i})$ has zero normal flux through (a part of) $\partial\mathcal{T}_{\mathbf{a}}$ together with (4.7) on $\partial\mathcal{T}_{\mathbf{a}} \cap \partial\Omega$ when $\mathbf{a} \in \mathcal{V}_h^{\mathrm{ext}}$, the fact that $\mathbf{d}_{\mathbf{a}}^{k,i} \in \mathbf{RTN}_l^{\mathrm{N},0}(\mathcal{T}_{\mathbf{a}})$, (4.7), and the properties (2.17) of the projection operator $\mathbf{I}_l^{\mathrm{RTN}}$ yield

$$\begin{aligned}
&-\sum_{K'\in\mathcal{T}_{\mathbf{a}}} (\nabla\cdot(\mathbf{d}_{\mathbf{a}}^{k,i} + \mathbf{I}_l^{\mathrm{RTN}}(\psi_{\mathbf{a}}\overline{\boldsymbol{\sigma}}_h^{k,i})), m_h)_{K'} + \sum_{e\in\mathcal{E}_h^{\mathrm{int}},\, e\cap\mathbf{a}\neq\emptyset} \langle [\![\mathbf{I}_l^{\mathrm{RTN}}(\psi_{\mathbf{a}}\overline{\boldsymbol{\sigma}}_h^{k,i})\cdot\mathbf{n}_e]\!], m_h\rangle_e \\
&= -\sum_{K'\in\mathcal{T}_{\mathbf{a}}} (\nabla\cdot(\mathbf{d}_{\mathbf{a}}^{k,i} + \psi_{\mathbf{a}}\overline{\boldsymbol{\sigma}}_h^{k,i}), \Pi_l(m_h))_{K'} + \sum_{e\in\mathcal{E}_h^{\mathrm{int}},\, e\cap\mathbf{a}\neq\emptyset} \langle [\![\psi_{\mathbf{a}}\overline{\boldsymbol{\sigma}}_h^{k,i}\cdot\mathbf{n}_e]\!], \Pi_l(m_h)\rangle_e
\end{aligned}$$

that we denote as $I + II$. Employing the second lines of the problems of Definition 4.2 (recall that we can take $\phi_h \in \mathbb{P}_l(\mathcal{T}_{\mathbf{a}})$), the first term I above can be developed as

$$\begin{aligned}
&-\sum_{K'\in\mathcal{T}_{\mathbf{a}}} (\nabla\cdot(\psi_{\mathbf{a}}\overline{\boldsymbol{\sigma}}_h^{k,i} + f\psi_{\mathbf{a}} - \overline{\boldsymbol{\sigma}}_h^{k,i}\cdot\nabla\psi_{\mathbf{a}} - \bar{R}_{\mathbf{a}}^{k,i}|\mathcal{T}_{\mathbf{a}}|^{-1}), \Pi_l(m_h))_{K'} \\
&= -\sum_{K'\in\mathcal{T}_{\mathbf{a}}} (\psi_{\mathbf{a}}(\nabla\cdot\overline{\boldsymbol{\sigma}}_h^{k,i} + f) - \bar{R}_{\mathbf{a}}^{k,i}|\mathcal{T}_{\mathbf{a}}|^{-1}, \Pi_l(m_h))_{K'} \\
&\leq \left\{\sum_{K'\in\mathcal{T}_{\mathbf{a}}} h_{K'}^{-p}\|m_h\|_{p,K'}^p\right\}^{\frac{1}{p}} \left\{\sum_{K'\in\mathcal{T}_{\mathbf{a}}} h_{K'}^q(\|f + \nabla\cdot\overline{\boldsymbol{\sigma}}_h^{k,i}\|_{q,K'} + \|\bar{R}_{\mathbf{a}}^{k,i}|\mathcal{T}_{\mathbf{a}}|^{-1}\|_{q,K'})^q\right\}^{\frac{1}{q}} \\
&\lesssim h_{\mathcal{T}_{\mathbf{a}}}^{-1}\|m_h\|_{p,\mathcal{T}_{\mathbf{a}}} \left(\left\{\sum_{K'\in\mathcal{T}_{\mathbf{a}}} h_{K'}^q\|f + \nabla\cdot\overline{\boldsymbol{\sigma}}_h^{k,i}\|_{q,K'}^q\right\}^{\frac{1}{q}} + |\bar{R}_{\mathbf{a}}^{k,i}||\mathcal{T}_{\mathbf{a}}|^{-1+\frac{1}{q}}h_{\mathcal{T}_{\mathbf{a}}}\right),
\end{aligned}$$

where we have also used the Hölder inequality, the stability of the Π_l-projection, and the fact that $\|\psi_{\mathbf{a}}\|_{\infty,\mathcal{T}_{\mathbf{a}}} = 1$. Finally, for any interior vertex $\mathbf{a}$, we get from (4.6), the Green theorem, the Hölder inequality, and the p-robust inverse inequality $\|\psi_{\mathbf{a}}\|_{p,e} \lesssim h_e^{-\frac{1}{p}}\|\psi_{\mathbf{a}}\|_{p,K'}$, $e \in \mathcal{E}_{K'}$, see [18, Section A.4], that the term $\bar{R}_{\mathbf{a}}^{k,i}$ can be developed as

$$\begin{aligned}
&\sum_{K'\in\mathcal{T}_{\mathbf{a}}} (f + \nabla\cdot\overline{\boldsymbol{\sigma}}_h^{k,i}, \psi_{\mathbf{a}})_{K'} - \sum_{e\in\mathcal{E}_h^{\mathrm{int}},\, e\cap\mathbf{a}\neq\emptyset} \langle [\![\overline{\boldsymbol{\sigma}}_h^{k,i}\cdot\mathbf{n}_e]\!], \psi_{\mathbf{a}}\rangle_e \\
&\lesssim \left(\left\{\sum_{K'\in\mathcal{T}_{\mathbf{a}}} h_{K'}^q\|f + \nabla\cdot\overline{\boldsymbol{\sigma}}_h^{k,i}\|_{q,K'}^q\right\}^{\frac{1}{q}} + \left\{\sum_{e\in\mathcal{E}_h^{\mathrm{int}},\, e\cap\mathbf{a}\neq\emptyset} h_e\|[\![\overline{\boldsymbol{\sigma}}_h^{k,i}\cdot\mathbf{n}_e]\!]\|_{q,e}^q\right\}^{\frac{1}{q}}\right) h_{\mathcal{T}_{\mathbf{a}}}^{-1}|\mathcal{T}_{\mathbf{a}}|^{\frac{1}{p}}.
\end{aligned}$$

Using the p-robust discrete Poincaré/Friedrichs inequality $\|m_h\|_{p,\mathcal{T}_\mathbf{a}} \lesssim h_{\mathcal{T}_\mathbf{a}} \|\nabla m_h\|_{p,\mathcal{T}_\mathbf{a}}$ from [18, Section A.4] and the triangle inequality for separating the data oscillation terms $\eta_{\mathrm{osc},K}^{k,i}$, we conclude that $I \le \eta_{\sharp,\mathfrak{T}_K}^{k,i} + \eta_{\mathrm{osc},\mathfrak{T}_K}^{k,i}$. Proceeding similarly for the jump term II (with the above treatment of $\psi_\mathbf{a}$ and Π_l) yields the desired result. □

5 Interior penalty discontinuous Galerkin (IPDG) for quasi-linear diffusion

We treat here the discretization of the quasi-linear diffusion problem (so that $p = q = 2$ and $\boldsymbol{\sigma}(v,\boldsymbol{\xi}) = \underline{\mathbf{A}}(v)\boldsymbol{\xi}$) by the IPDG method. We focus on flux reconstruction by full prescription, which is the easiest in practice, extending the work of Kim [28] and [19] on discretization errors. Reconstruction by local mixed problems with prescription can be achieved by proceeding as in [20]; reconstruction without prescription can also be considered by proceeding as in Section 6.4.

5.1 Discretization

Let $V_h := \mathbb{P}_m(\mathcal{T}_h)$, $m \ge 1$. The IPDG discretization of problem (1.2) reads: find $u_h \in V_h$ such that, for all $v_h \in V_h$,

$$\begin{aligned}(\boldsymbol{\sigma}(u_h,\nabla u_h),\nabla v_h) - \sum_{e\in\mathcal{E}_h}\big\{\langle\{\!\{\boldsymbol{\sigma}(u_h,\nabla u_h)\}\!\}\cdot\mathbf{n}_e,[\![v_h]\!]\rangle_e\\ + \theta\langle\{\!\{\underline{\mathbf{A}}(u_h)\nabla v_h\}\!\}\cdot\mathbf{n}_e,[\![u_h]\!]\rangle_e\big\} + \sum_{e\in\mathcal{E}_h}\langle\bar\alpha_e h_e^{-1}[\![u_h]\!],[\![v_h]\!]\rangle_e = (f,v_h),\end{aligned} \tag{5.1}$$

with $\theta \in \{-1,0,1\}$ and $\bar\alpha_e := \|\underline{\mathbf{A}}\|_{L^\infty(\mathbb{R})}\chi_e$ where χ_e is a large enough positive parameter. The average operator $\{\!\{\cdot\}\!\}$ yields the mean value of the traces from adjacent mesh elements on interfaces and the actual trace on boundary faces. Testing (5.1) against the basis functions in V_h gives rise to the nonlinear algebraic system (2.3); these basis functions are denoted $\psi_{K,j}$, for all $K \in \mathcal{T}_h$ and all $j \in \mathcal{C}_K := \{1,\dots,\dim(\mathbb{P}_m(K))\}$.

5.2 Linearization

Let $u_h^0 \in V_h$, fixing the initial vector U^0 in Algorithm 2.1. The linearization of (5.1), for $k \ge 1$, reads: find $u_h^k \in V_h$ such that, for all $K \in \mathcal{T}_h$ and all $j \in \mathcal{C}_K$,

$$\begin{aligned}(\boldsymbol{\sigma}^{k-1}(u_h^k,\nabla u_h^k),\nabla\psi_{K,j}) - \sum_{e\in\mathcal{E}_h}\big\{\langle\{\!\{\boldsymbol{\sigma}^{k-1}(u_h^k,\nabla u_h^k)\}\!\}\cdot\mathbf{n}_e,[\![\psi_{K,j}]\!]\rangle_e\\ + \theta\langle\{\!\{\underline{\mathbf{A}}^{k-1}(u_h^k)\nabla\psi_{K,j}\}\!\}\cdot\mathbf{n}_e,[\![u_h^k]\!]\rangle_e\big\} + \sum_{e\in\mathcal{E}_h}\langle\bar\alpha_e h_e^{-1}[\![u_h^k]\!],[\![\psi_{K,j}]\!]\rangle_e = (f,\psi_{K,j}),\end{aligned} \tag{5.2}$$

which is the functional form of (2.4). The fixed point linearization corresponds to $\boldsymbol{\sigma}^{k-1}(v,\boldsymbol{\xi}) := \underline{\mathbf{A}}(u_h^{k-1})\boldsymbol{\xi}$ and $\underline{\mathbf{A}}^{k-1}(v) := \underline{\mathbf{A}}(u_h^{k-1})$, and the Newton linearization to

$$\boldsymbol{\sigma}^{k-1}(v,\boldsymbol{\xi}) := \underline{\mathbf{A}}(u_h^{k-1})\boldsymbol{\xi} + (v-u_h^{k-1})\partial_v\underline{\mathbf{A}}(u_h^{k-1})\nabla u_h^{k-1}, \tag{5.3a}$$

$$\underline{\mathbf{A}}^{k-1}(v) := \underline{\mathbf{A}}(u_h^{k-1}) + \partial_v\underline{\mathbf{A}}(u_h^{k-1})(v-u_h^{k-1}). \tag{5.3b}$$

5.3 Algebraic solution

On the i-th step, $i \ge 1$, of an iterative linear solver for the algebraic system (2.4), we obtain the system (2.5) with algebraic residual vector $R^{k,i} = \{R_{K,j}^{k,i}\}_{K\in\mathcal{T}_h,\,j\in\mathcal{C}_K}$. The functional form of (2.5) is: find $u_h^{k,i} \in V_h$ such that, for all $K \in \mathcal{T}_h$ and all $j \in \mathcal{C}_K$,

$$\begin{aligned}(\boldsymbol{\sigma}^{k-1}(u_h^{k,i},\nabla u_h^{k,i}),\nabla\psi_{K,j}) - \sum_{e\in\mathcal{E}_h}\big\{\langle\{\!\{\boldsymbol{\sigma}^{k-1}(u_h^{k,i},\nabla u_h^{k,i})\}\!\}\cdot\mathbf{n}_e,[\![\psi_{K,j}]\!]\rangle_e\\ +\theta\langle\{\!\{\underline{\mathbf{A}}^{k-1}(u_h^{k,i})\nabla\psi_{K,j}\}\!\}\cdot\mathbf{n}_e,[\![u_h^{k,i}]\!]\rangle_e\big\}+\sum_{e\in\mathcal{E}_h}\langle\bar\alpha_e h_e^{-1}[\![u_h^{k,i}]\!],[\![\psi_{K,j}]\!]\rangle_e = (f,\psi_{K,j}) - R_{K,j}^{k,i}.\end{aligned}$$

5.4 Flux reconstruction by full prescription

We construct $\mathbf{d}_h^{k,i}$ and $\mathbf{l}_h^{k,i}$ in the space $\mathbf{RTN}_l(\mathcal{T}_h)$ with $l := m-1$ or $l := m$. For all $e \in \mathcal{E}_h$, we set $w_e := \frac{1}{2}$ if $e \in \mathcal{E}_h^{\rm int}$ and $w_e := 1$ if $e \in \mathcal{E}_h^{\rm ext}$.

Definition 5.1 (Construction of $(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i})$)**.** *The function $(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i})$ is defined in $\mathbf{RTN}_l(\mathcal{T}_h)$ such that, for all $K \in \mathcal{T}_h$ and all $e \in \mathcal{E}_K$,*

$$\langle(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i})\cdot\mathbf{n}_e, q_h\rangle_e := \langle -\{\!\!\{\boldsymbol{\sigma}^{k-1}(u_h^{k,i},\nabla u_h^{k,i})\}\!\!\}\cdot\mathbf{n}_e + \bar\alpha_e h_e^{-1}[\![u_h^{k,i}]\!], q_h\rangle_e,$$
$$(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i}, \mathbf{r}_h)_K := -(\boldsymbol{\sigma}^{k-1}(u_h^{k,i},\nabla u_h^{k,i}), \mathbf{r}_h)_K + \theta\sum_{e\in\mathcal{E}_K} w_e\langle\underline{\mathbf{A}}^{k-1}(u_h^{k,i})\mathbf{r}_h\cdot\mathbf{n}_e, [\![u_h^{k,i}]\!]\rangle_e,$$

for all $q_h \in \mathbb{P}_l(e)$ and all $\mathbf{r}_h \in [\mathbb{P}_{l-1}(K)]^d$.

Definition 5.2 (Construction of $\mathbf{d}_h^{k,i}$)**.** *The function $\mathbf{d}_h^{k,i}$ is in $\mathbf{RTN}_l(\mathcal{T}_h)$ and is defined using the prescription of Definition 5.1 with $\boldsymbol{\sigma}(u_h^{k,i},\nabla u_h^{k,i})$ in place of $\boldsymbol{\sigma}^{k-1}(u_h^{k,i},\nabla u_h^{k,i})$ and $\underline{\mathbf{A}}(u_h^{k,i})$ in place of $\underline{\mathbf{A}}^{k-1}(u_h^{k,i})$.*

Definition 5.3 (Error measure, data oscillation, quadrature, and algebraic remainder)**.** *Use $u_h^{k,i}$ and $\mathbf{g}_h^{k,i} := \nabla u_h^{k,i}$ in the error measure (2.1) and set $f_h := \Pi_l f$, $\overline{\boldsymbol{\sigma}}_h^{k,i} := \mathbf{I}_l^{\rm RTN}(\boldsymbol{\sigma}(u_h^{k,i},\nabla u_h^{k,i}))$, and $r_h^{k,i} \in \mathbb{P}_m(\mathcal{T}_h)$ with $(r_h^{k,i},\psi_{K,j})_K = R_{K,j}^{k,i}$ for all $K \in \mathcal{T}_h$ and all $j \in \mathcal{C}_K$.*

We now verify the assumptions of Section 2.5:

Lemma 5.4 (Quasi-equilibration)**.** *Assumption 2.3 holds.*

Proof. Direct verification by proceeding as in [19, 28], see also [17, Section 5.5]. □

Lemma 5.5 (Local approximation and convergence)**.** *Assumption 2.4 holds using weights $\alpha_e := \bar\alpha_e^2$ and exponent $s := p$ in the nonconformity estimator.*

Proof. The requirement on $\mathbf{l}_h^{k,i}$ is obvious from Definitions 5.1 and 5.2. Turning to $\mathbf{d}_h^{k,i}$, we observe that, for all $K \in \mathcal{T}_h$ and all $e \in \mathcal{E}_K$, there holds

$$\langle(\mathbf{d}_h^{k,i}+\overline{\boldsymbol{\sigma}}_h^{k,i})\cdot\mathbf{n}_e, q_h\rangle_e = (1-w_e)\langle[\![\overline{\boldsymbol{\sigma}}_h^{k,i}]\!]\cdot\mathbf{n}_e + \bar\alpha_e h_e^{-1}[\![u_h^{k,i}]\!], q_h\rangle_e, \tag{5.4a}$$
$$(\mathbf{d}_h^{k,i}+\overline{\boldsymbol{\sigma}}_h^{k,i}, \mathbf{r}_h)_K = \theta\sum_{e\in\mathcal{E}_K} w_e\langle\underline{\mathbf{A}}(u_h^{k,i})\mathbf{r}_h\cdot\mathbf{n}_e, [\![u_h^{k,i}]\!]\rangle_e, \tag{5.4b}$$

for all $q_h \in \mathbb{P}_l(e)$ and all $\mathbf{r}_h \in [\mathbb{P}_{l-1}(K)]^d$. The assertion then follows from standard approximation properties in Raviart–Thomas–Nédélec spaces, see, e.g., [17, Section 5.5]. □

6 Discontinuous Galerkin with gradient reconstruction

We treat here the discretization of problem (1.2) by the discontinuous Galerkin method. A key ingredient, especially for the Leray–Lions setting, is the definition of a suitable discrete gradient. Regarding flux reconstruction, we introduce a new approach based on local mixed problems without prescription on patches of elements.

6.1 Discretization

Let $l \geq 0$ be an integer. For all $e \in \mathcal{E}_h$, we define the map $\ell_e : L^1(e) \to [\mathbb{P}_l(\mathcal{T}_h)]^d$ such that, for all $\phi \in L^1(e)$, $\ell_e(\phi)$ is the unique function in $[\mathbb{P}_l(\mathcal{T}_h)]^d$ such that, for all $\mathbf{v}_h \in [\mathbb{P}_l(\mathcal{T}_h)]^d$, $(\ell_e(\phi),\mathbf{v}_h) = \langle\{\!\!\{\mathbf{v}_h\}\!\!\}\cdot\mathbf{n}_e, \phi\rangle_e$. The vector-valued, piecewise polynomial function $\ell_e(\phi)$ is supported in $\mathcal{T}_e$ (the patch of elements sharing the face e) and is colinear to $\mathbf{n}_e$. Then, for a function $v \in V(\mathcal{T}_h)$, we define its discrete gradient $\nabla_h v \in [L^p(\Omega)]^d$ (see [17, Section 4.2] and the references therein) as

$$\nabla_h v := \nabla v - \mathtt{L}_h([\![v]\!]), \qquad \mathtt{L}_h([\![v]\!]) := \sum_{e\in\mathcal{E}_h}\ell_e([\![v]\!]). \tag{6.1}$$

We observe that $\mathtt{L}_h([\![v]\!])$ is a (piecewise polynomial) correction to the broken gradient ∇v based on the liftings of the jumps. The discrete gradient is an important tool in the design of discontinuous Galerkin methods for nonlinear problems, see Buffa and Ortner [7] and [8] for the p-Laplacian and [16] for the incompressible Navier–Stokes equations.

Let $V_h := \mathbb{P}_m(\mathcal{T}_h)$, $m \geq 1$. We consider here the following gradient reconstruction discontinuous Galerkin method: find $u_h \in V_h$ such that

$$(\boldsymbol{\sigma}(u_h, \nabla_h u_h), \nabla_h v_h) + \sum_{e\in\mathcal{E}_h} \langle s_e([\![u_h]\!]), [\![v_h]\!]\rangle_e = (f, v_h) \qquad \forall v_h \in V_h, \tag{6.2}$$

with the stabilization operator $s_e : L^p(e) \to L^q(e)$ for all $e \in \mathcal{E}_h$ such that, for all $v \in L^p(e)$, $s_e(v) = \bar\alpha_e h_e^{1-p}|v|^{p-2}v$ with a positive parameter $\bar\alpha_e$. Testing (6.2) against the basis functions in V_h gives rise to the nonlinear algebraic system (2.3).

Remark 6.1 (Stencil reduction and link with IPDG)**.** *The discretization stencil resulting from* (6.2) *includes neighbors and neighbors of neighbors in the sense of faces. This stencil can be reduced to the more classical IPDG stencil only including face neighbors by adding to the left-hand side of* (6.2) *the form*

$$-(\boldsymbol{\sigma}(u_h, \nabla_h u_h) - \boldsymbol{\sigma}(u_h, \nabla u_h), \nabla_h v_h - \nabla v_h) \tag{6.3}$$

yielding, for all $v_h \in V_h$*,*

$$(\boldsymbol{\sigma}(u_h, \nabla_h u_h), \nabla v_h) - (\boldsymbol{\sigma}(u_h, \nabla u_h), \mathtt{L}_h([\![v_h]\!])) + \sum_{e\in\mathcal{E}_h} \langle s_e([\![u_h]\!]), [\![v_h]\!]\rangle_e = (f, v_h).$$

Since the form (6.3) *is negative, this modification requires the penalty term to be strong enough to control it. For the quasi-linear diffusion problem, this is classically achieved by taking* $\bar\alpha_e = \|\underline{\mathbf{A}}\|_{L^\infty(\mathbb{R})}\chi_e$ *and* χ_e *large enough* [17, Section 4.3]*. Moreover, this modification leads to an IPDG formulation of the type* (5.1) *with* $\theta = 1$ *and, for all* $v_h \in V_h$*,*

$$\begin{aligned}(\boldsymbol{\sigma}(u_h, \nabla u_h), \nabla v_h) - \sum_{e\in\mathcal{E}_h} \big\{ &\langle \{\!\{ \mathbf{I}_l^{\mathrm{RTN}}(\boldsymbol{\sigma}(u_h, \nabla u_h))\}\!\}\cdot\mathbf{n}_e, [\![v_h]\!]\rangle_e \\ &+ \langle \{\!\{ \mathbf{I}_l^{\mathrm{RTN}}(\underline{\mathbf{A}}(u_h)\nabla v_h)\}\!\}\cdot\mathbf{n}_e, [\![u_h]\!]\rangle_e \big\} + \sum_{e\in\mathcal{E}_h} \langle \bar\alpha_e h_e^{-1}[\![u_h]\!], [\![v_h]\!]\rangle_e = (f, v_h).\end{aligned}$$

6.2 Linearization

Let $u_h^0 \in V_h$, fixing the initial vector U^0 in Algorithm 2.1. The linearization of (6.2), for $k \geq 1$, reads: find $u_h^k \in V_h$ such that, for all $K \in \mathcal{T}_h$ and all $j \in \mathcal{C}_K := \{1, \ldots, \dim(\mathbb{P}_m(K))\}$,

$$(\boldsymbol{\sigma}^{k-1}(u_h^k, \nabla_h u_h^k), \nabla_h \psi_{K,j}) + \sum_{e\in\mathcal{E}_h} \langle s_e^{k-1}([\![u_h^k]\!]), [\![\psi_{K,j}]\!]\rangle_e = (f, \psi_{K,j}), \tag{6.4}$$

which is the functional form of the algebraic system (2.4). In the fixed-point linearization, $\boldsymbol{\sigma}^{k-1}(v, \boldsymbol{\xi})$ is defined by (3.3) with $\nabla_h u_h^{k-1}$ in place of ∇u_h^{k-1}, while $s_e^{k-1}(v) := \bar\alpha_e h_e^{1-p}|[\![u_h^{k-1}]\!]|^{p-2}v$. In the Newton linearization, $\boldsymbol{\sigma}^{k-1}(v, \boldsymbol{\xi})$ is defined by (3.4) with $\nabla_h u_h^{k-1}$ in place of ∇u_h^{k-1}, while $s_e^{k-1}(v) := \bar\alpha_e h_e^{1-p}|[\![u_h^{k-1}]\!]|^{p-2}((p-1)v - (p-2)[\![u_h^{k-1}]\!])$.

6.3 Algebraic solution

On the i-th step, $i \geq 1$, of an iterative linear solver for the algebraic system (2.4), we obtain the system (2.5) with algebraic residual vector $R^{k,i} = \{R_{K,j}^{k,i}\}_{K\in\mathcal{T}_h,\, j\in\mathcal{C}_K}$. The functional form of (2.5) is: find $u_h^{k,i} \in V_h$ such that, for all $K \in \mathcal{T}_h$ and all $j \in \mathcal{C}_K$,

$$(\boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla_h u_h^{k,i}), \nabla_h \psi_{K,j}) + \sum_{e\in\mathcal{E}_h} \langle s_e^{k-1}([\![u_h^{k,i}]\!]), [\![\psi_{K,j}]\!]\rangle_e = (f, \psi_{K,j}) - R_{K,j}^{k,i}. \tag{6.5}$$

6.4 Flux reconstruction by local mixed problems without prescription

We proceed as in Section 4.4 hinging on the hat basis functions $\psi_{\mathbf{a}} \in \mathbb{P}_1(\mathcal{T}_h) \cap C^0(\Omega)$ associated with the vertices $\mathbf{a} \in \mathcal{V}_h$. This in particular allows us to eliminate the nonlinear jump terms in the local flux expressions, compare with (5.4). Since $m \geq 1$, there holds $\psi_{\mathbf{a}} \in V_h$, so that there are coefficients $\beta_{K,j}$ such that $\psi_{\mathbf{a}} = \sum_{K\in\mathcal{T}_{\mathbf{a}}} \sum_{j\in\mathcal{C}_K} \beta_{K,j}\psi_{K,j}$, recalling that $\mathcal{T}_{\mathbf{a}}$ is the patch of elements of $\mathcal{T}_h$ sharing $\mathbf{a}$. We can then distribute the components of the algebraic residual vector onto vertices by setting $R_{\mathbf{a}}^{k,i} := \sum_{K\in\mathcal{T}_{\mathbf{a}}} \sum_{j\in\mathcal{C}_K} \beta_{K,j} R_{K,j}^{k,i}$ for all $\mathbf{a} \in \mathcal{V}_h^{\rm int}$, and $R_{\mathbf{a}}^{k,i} := 0$ for all $\mathbf{a} \in \mathcal{V}_h^{\rm ext}$.

We construct $\mathbf{d}_h^{k,i}$ and $\mathbf{l}_h^{k,i}$ in the space $\mathbf{RTN}_l(\mathcal{T}_h)$ with $l := m-1$ or $l := m$. We use the notation from Section 4.4.

Definition 6.2 (Construction of $(\mathbf{d}_h^{k,i} + \mathbf{l}_h^{k,i})$)**.** *We define* $(\mathbf{d}_h^{k,i} + \mathbf{l}_h^{k,i}) \in \mathbf{RTN}_l(\mathcal{T}_h)$ *using Definition 4.1 with* $\boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla_h u_h^{k,i})$ *in place of* $\boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla u_h^{k,i})$.

In the local mixed problems considered in Definition 4.1, we can take $\phi_h \in \mathbb{P}_l(\mathcal{T}_{\mathbf{a}})$ since multiplying (6.5) by the coefficients $\beta_{K,j}$, summing over all $K \in \mathcal{T}_{\mathbf{a}}$ and all $j \in \mathcal{C}_K$, using the definition of $R_{\mathbf{a}}^{k,i}$, and the fact that $[\![\psi_{\mathbf{a}}]\!] = 0$, yields, for all $\mathbf{a} \in \mathcal{V}_h^{\rm int}$, the Neumann compatibility condition

$$(\boldsymbol{\sigma}^{k-1}(u_h^{k,i}, \nabla_h u_h^{k,i}), \nabla\psi_{\mathbf{a}})_{\mathcal{T}_{\mathbf{a}}} = (f, \psi_{\mathbf{a}})_{\mathcal{T}_{\mathbf{a}}} - R_{\mathbf{a}}^{k,i}. \tag{6.6}$$

We proceed similarly for the construction of $\mathbf{d}_h^{k,i}$, setting, for all $\mathbf{a} \in \mathcal{V}_h^{\rm int}$, $\bar{R}_{\mathbf{a}}^{k,i} := (f, \psi_{\mathbf{a}})_{\mathcal{T}_{\mathbf{a}}} - (\boldsymbol{\sigma}(u_h^{k,i}, \nabla_h u_h^{k,i}), \nabla\psi_{\mathbf{a}})_{\mathcal{T}_{\mathbf{a}}}$ and, for all $\mathbf{a} \in \mathcal{V}_h^{\rm ext}$, $\bar{R}_{\mathbf{a}}^{k,i} := 0$. This yields:

Definition 6.3 (Construction of $\mathbf{d}_h^{k,i}$)**.** *We define* $\mathbf{d}_h^{k,i} \in \mathbf{RTN}_l(\mathcal{T}_h)$ *using Definition 4.2 with* $\boldsymbol{\sigma}(u_h^{k,i}, \nabla_h u_h^{k,i})$ *in place of* $\boldsymbol{\sigma}(u_h^{k,i}, \nabla u_h^{k,i})$.

Definition 6.4 (Error measure, data oscillation, quadrature, and algebraic remainder)**.** *Use* $u_h^{k,i}$ *and* $\mathbf{g}_h^{k,i} := \nabla_h u_h^{k,i}$ *in the error measure (2.1) and set* $f_h := \Pi_l f$, $\overline{\boldsymbol{\sigma}}_h^{k,i} := \mathbf{\Pi}_l(\boldsymbol{\sigma}(u_h^{k,i}, \nabla_h u_h^{k,i}))$, *and* $r_h^{k,i}|_K := \sum_{\mathbf{a}\in\mathcal{V}_K} |\mathcal{T}_{\mathbf{a}}|^{-1} R_{\mathbf{a}}^{k,i}$ *for all* $K \in \mathcal{T}_h$, *recalling that* $\mathcal{V}_K$ *collects the vertices of the element* K.

Owing to the results of Section 4.4, which exactly apply here as well, the assumptions of Section 2.5 are satisfied:

Lemma 6.5 (Quasi-equilibration)**.** *Assumption 2.3 holds.*

Lemma 6.6 (Local approximation and convergence)**.** *Assumption 2.4 holds.*

7 Cell-centered finite volumes and lowest-order mixed finite elements

We consider here the discretization of problem (1.2) by a general cell-centered finite volume method, cf. Eymard *et al.* [22]. We also treat in the last subsection the discretization by lowest-order mixed finite elements using the links between the two methods, cf. [37, 40]. Very few results are available concerning the a posteriori error analysis of such schemes for problem (1.2). We mention the Ph.D. thesis of Ahn Ha Le [29] which extends the analysis of [18] to cell-centered finite volume discretizations of quasi-linear diffusion problems with fixed point linearization.

7.1 Discretization

Let $V_h := \mathbb{P}_0(\mathcal{T}_h)$. Fix an element $K \in \mathcal{T}_h$ and a face $e \in \mathcal{E}_K$. We denote by $\sigma_{K,e} : \mathbb{P}_0(\mathcal{T}_h) \to \mathbb{R}$ the finite volume flux function, which maps a piecewise constant function $\bar{v}_h \in V_h$ to the normal flux through the face e, $\sigma_{K,e}(\bar{v}_h)$. We do not need the specific form of the flux functions $\sigma_{K,e}$, except that conservativity be satisfied in the form $\sigma_{K,e}(\bar{v}_h) = -\sigma_{K',e}(\bar{v}_h)$ for any function $\bar{v}_h \in V_h$ and any interface $e \in \mathcal{E}_h^{\rm int}$ such that $e = \partial K \cap \partial K'$. A general cell-centered finite volume method for the problem (1.2) reads: find $\bar{u}_h \in V_h$ such that

$$\sum_{e\in\mathcal{E}_K} \sigma_{K,e}(\bar{u}_h) = (f, 1)_K \qquad \forall K \in \mathcal{T}_h. \tag{7.1}$$

This gives rise to the nonlinear algebraic system (2.3).

7.2 Linearization

Let $\bar{u}_h^0 \in V_h$, fixing the initial vector U^0 in Algorithm 2.1. The linearization of (7.1), for $k \geq 1$, reads: find $\bar{u}_h^k \in V_h$ such that

$$\sum_{e \in \mathcal{E}_K} \sigma_{K,e}^{k-1}(\bar{u}_h^k) = (f,1)_K \qquad \forall K \in \mathcal{T}_h, \tag{7.2}$$

which is the functional form of the algebraic system (2.4). Here, $\sigma_{K,e}^{k-1} : V_h \to \mathbb{R}$ is the finite volume flux function on the k-th linearization step. We again suppose conservativity, i.e., $\sigma_{K,e}^{k-1}(\bar{v}_h) = -\sigma_{K',e}^{k-1}(\bar{v}_h)$ for any $\bar{v}_h \in V_h$ and $e = \partial K \cap \partial K' \in \mathcal{E}_h^{\mathrm{int}}$. It is not possible to specify the fixed point linearization directly from (7.1), as it depends on the actual form of $\sigma_{K,e}$. For the Newton linearization, $\sigma_{K,e}^{k-1}$ is such that

$$\sigma_{K,e}^{k-1}(\bar{v}_h) := \sigma_{K,e}(\bar{u}_h^{k-1}) + \sum_{K' \in \mathcal{T}_h} \frac{\partial \sigma_{K,e}}{\partial \bar{u}_h|_{K'}}(\bar{u}_h^{k-1})(\bar{v}_h|_{K'} - \bar{u}_h^{k-1}|_{K'}). \tag{7.3}$$

As an example, we detail the linearized flux function $\sigma_{K,e}^{k-1}$ for a two-point finite volume scheme. Let $d = 2$ and assume that $\mathcal{T}_h$ is strictly Delaunay, so that the circumcircle of each triangle does not contain any other triangle vertex, and each circumcenter of a boundary triangle is inside Ω. Consider the quasi-linear diffusion setting with a scalar-valued function $a(\mathbf{x}, v)$ (in place of the tensor-valued function $\underline{\mathbf{A}}(\mathbf{x}, v)$). Let $\mathbf{x}_K^\circ$ stand for the circumcenter of the triangle $K \in \mathcal{T}_h$ and $\mathbf{x}_e$ for the center of the edge $e \in \mathcal{E}_h^{\mathrm{ext}}$. We use the shorthand notation $a_K(\cdot)$ in place of $a(\mathbf{x}_K^\circ, \cdot)$ and $\bar{v}_K$ in place of $\bar{v}_h|_K$ for any function $\bar{v}_h \in V_h$. Then, a two-point finite volume scheme for the quasi-linear diffusion problem takes the form (7.1) with

$$\sigma_{K,e}(\bar{u}_h) := \frac{k_e}{2}\Big\{a_K(\bar{u}_K) + a_{K'}(\bar{u}_{K'})\Big\}(\bar{u}_K - \bar{u}_{K'}) \qquad \forall e = \partial K \cap \partial K' \in \mathcal{E}_h^{\mathrm{int}}, \tag{7.4a}$$

$$\sigma_{K,e}(\bar{u}_h) := k_e a_K(\bar{u}_K)\bar{u}_K \qquad \forall e = \partial K \cap \partial\Omega \in \mathcal{E}_h^{\mathrm{ext}}, \tag{7.4b}$$

where $k_e := \frac{|e|}{|\mathbf{x}_K^\circ - \mathbf{x}_{K'}^\circ|}$ in (7.4a) and $k_e := \frac{|e|}{|\mathbf{x}_K^\circ - \mathbf{x}_e|}$ in (7.4b). The Newton linearization leads to, for all $K \in \mathcal{T}_h$ and all $e = \partial K \cap \partial K' \in \mathcal{E}_h^{\mathrm{int}}$,

$$\begin{aligned}\sigma_{K,e}^{k-1}(\bar{v}_h) := {}& \frac{k_e}{2}\Big\{a_K(\bar{u}_K^{k-1}) + a_{K'}(\bar{u}_{K'}^{k-1})\Big\}(\bar{v}_K - \bar{v}_{K'}) \\ & + \frac{k_e}{2}\Big\{a_K'(\bar{u}_K^{k-1})(\bar{v}_K - \bar{u}_K^{k-1}) + a_{K'}'(\bar{u}_{K'}^{k-1})(\bar{v}_{K'} - \bar{u}_{K'}^{k-1})\Big\}(\bar{u}_K^{k-1} - \bar{u}_{K'}^{k-1}),\end{aligned} \tag{7.5}$$

and, for all $e = \partial K \cap \partial\Omega \in \mathcal{E}_h^{\mathrm{ext}}$,

$$\sigma_{K,e}^{k-1}(\bar{v}_h) := k_e a_K(\bar{u}_K^{k-1})\bar{v}_K + k_e a_K'(\bar{u}_K^{k-1})(\bar{v}_K - \bar{u}_K^{k-1})\bar{u}_K^{k-1}. \tag{7.6}$$

Moreover, the fixed point linearization is derived from (7.5)–(7.6) by omitting the terms with the derivative of a.

7.3 Algebraic solution

On the i-th step, $i \geq 1$, of an iterative linear solver for the algebraic system (2.4), we obtain the algebraic residual vector $R^{k,i}$ in (2.5) with $R^{k,i} = \{R_K^{k,i}\}_{K \in \mathcal{T}_h}$. The functional form of (2.5) is: find $\bar{u}_h^{k,i} \in V_h$ such that

$$\sum_{e \in \mathcal{E}_K} \sigma_{K,e}^{k-1}(\bar{u}_h^{k,i}) = (f,1)_K - R_K^{k,i} \qquad \forall K \in \mathcal{T}_h. \tag{7.7}$$

7.4 Flux reconstruction by full prescription

Following Eymard *et al.* [23] and [38], we construct $\mathbf{d}_h^{k,i}$ and $\mathbf{l}_h^{k,i}$ in the space $\mathbf{RTN}_0(\mathcal{T}_h)$.

Definition 7.1 (Construction of $(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i})$)**.** *The function* $(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i})$ *is defined in* $\mathbf{RTN}_0(\mathcal{T}_h)$ *such that, for all* $K\in\mathcal{T}_h$ *and all* $e\in\mathcal{E}_K$*,*

$$\langle(\mathbf{d}_h^{k,i}+\mathbf{l}_h^{k,i})\cdot\mathbf{n}_K,1\rangle_e=\sigma_{K,e}^{k-1}(\bar u_h^{k,i}). \tag{7.8}$$

Definition 7.2 (Construction of $\mathbf{d}_h^{k,i}$)**.** *The function* $\mathbf{d}_h^{k,i}$ *is defined in* $\mathbf{RTN}_0(\mathcal{T}_h)$ *using Definition 7.1 with* $\sigma_{K,e}(\bar u_h^{k,i})$ *in place of* $\sigma_{K,e}^{k-1}(\bar u_h^{k,i})$*.*

The piecewise constant discrete potential $\bar u_h^{k,i}\in V_h$ has not enough regularity to be meaningful as an argument in the error measure (2.1), in particular regarding the size of its jumps. For this reason, following [38, 39], we introduce an elementwise postprocessing of $\bar u_h^{k,i}$, leading to a new discrete potential $u_h^{k,i}$ sitting in the richer polynomial space $\mathbb{P}_2(\mathcal{T}_h)$. The first step is to determine $\nabla u_h^{k,i}$ from $\mathbf{d}_h^{k,i}$. For simplicity, we assume that the $\boldsymbol{\xi}$-dependency of $\boldsymbol{\sigma}$ can be inverted, i.e., there is a function $\underline{\mathbf{B}}:\Omega\times\mathbb{R}\times\mathbb{R}^d\to\mathbb{R}^{d,d}$ such that, for all $(\mathbf{x},v,\boldsymbol{\xi},\boldsymbol{\tau})\in\Omega\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^d$,

$$\boldsymbol{\tau}=\underline{\mathbf{A}}(\mathbf{x},v,\boldsymbol{\xi})\boldsymbol{\xi}\iff\boldsymbol{\xi}=\underline{\mathbf{B}}(\mathbf{x},v,\boldsymbol{\tau})\boldsymbol{\tau}. \tag{7.9}$$

For the quasi-linear diffusion problem, there holds $\underline{\mathbf{B}}(\mathbf{x},v)=\underline{\mathbf{A}}(\mathbf{x},v)^{-1}$, while for the Leray–Lions problem in the p-Laplace setting, $\underline{\mathbf{B}}(\boldsymbol{\tau})=|\boldsymbol{\tau}|^{q-2}\underline{\mathbf{I}}$. Then, we set

$$\nabla u_h^{k,i}|_K:=\underline{\mathbf{B}}(\mathbf{x}_K,\bar u_h^{k,i}|_K,\mathbf{d}_h^{k,i}(\mathbf{x}_K))\mathbf{d}_h^{k,i}|_K\qquad\forall K\in\mathcal{T}_h, \tag{7.10}$$

where $\mathbf{x}_K$ denotes the barycenter or the circumcenter of K. Once $\nabla u_h^{k,i}$ is known, the second step is to determine a suitable integration constant in each element $K\in\mathcal{T}_h$. Possible choices are (depending on the finite volume scheme at hand) $(u_h^{k,i},1)_K/|K|:=\bar u_h^{k,i}|_K$ or $u_h^{k,i}(\mathbf{x}_K):=\bar u_h^{k,i}|_K$. This now fully defines $u_h^{k,i}\in\mathbb{P}_2(\mathcal{T}_h)$.

Definition 7.3 (Error measure, data oscillation, quadrature, and algebraic remainder)**.** *Use* $u_h^{k,i}$ *and* $\mathbf{g}_h^{k,i}:=\nabla u_h^{k,i}$ *in the error measure (2.1) and set* $f_h:=\Pi_0 f$*,* $\overline{\boldsymbol{\sigma}}_h^{k,i}:=\mathbf{d}_h^{k,i}$*, and* $r_h^{k,i}|_K:=|K|^{-1}R_K^{k,i}$ *for all* $K\in\mathcal{T}_h$*.*

Finally, the assumptions of Section 2.5 are readily verified:

Lemma 7.4 (Quasi-equilibration)**.** *Assumption 2.3 holds.*

Lemma 7.5 (Local approximation and convergence)**.** *Assumption 2.4 holds.*

7.5 Lowest-order mixed finite elements

We assume that the $\boldsymbol{\xi}$-dependency of $\boldsymbol{\sigma}$ can be inverted, see (7.9), and, omitting the $\mathbf{x}$-dependency, we set $\boldsymbol{\gamma}(v,\boldsymbol{\tau}):=\underline{\mathbf{B}}(v,\boldsymbol{\tau})\boldsymbol{\tau}$ for all $(v,\boldsymbol{\tau})\in\mathbb{R}\times\mathbb{R}^d$. Let $V_h:=\mathbb{P}_0(\mathcal{T}_h)$ and $\mathbf{V}_h:=\mathbf{RTN}_0(\mathcal{T}_h)$. The lowest-order Raviart–Thomas mixed finite element method to discretize problem (1.2) reads: find $(\boldsymbol{\sigma}_h,\bar u_h)\in\mathbf{V}_h\times V_h$ such that, for all $(\mathbf{v}_h,v_h)\in\mathbf{V}_h\times V_h$,

$$(\boldsymbol{\gamma}(\bar u_h,\boldsymbol{\sigma}_h),\mathbf{v}_h)-(\bar u_h,\nabla\cdot\mathbf{v}_h)=0, \tag{7.11a}$$

$$(\nabla\cdot\boldsymbol{\sigma}_h,v_h)=(f,v_h). \tag{7.11b}$$

This gives rise to the nonlinear algebraic system (2.3). The discrete problem (7.11) has been considered by Milner [32] in the quasi-linear diffusion setting and, e.g., by Creusé *et al.* [13] for the p-Laplacian.

Let $(\boldsymbol{\sigma}_h^0,\bar u_h^0)\in\mathbf{V}_h\times V_h$, fixing the initial vector U^0 in Algorithm 2.1. The linearization of (7.11), for $k\geq1$, reads: find $(\boldsymbol{\sigma}_h^k,\bar u_h^k)\in\mathbf{V}_h\times V_h$ such that, for all $(\mathbf{v}_h,v_h)\in\mathbf{V}_h\times V_h$,

$$(\boldsymbol{\gamma}^{k-1}(\bar u_h^k,\boldsymbol{\sigma}_h^k),\mathbf{v}_h)-(\bar u_h^k,\nabla\cdot\mathbf{v}_h)=0, \tag{7.12a}$$

$$(\nabla\cdot\boldsymbol{\sigma}_h^k,v_h)=(f,v_h), \tag{7.12b}$$

which is the functional form of the algebraic system (2.4). Two common ways to define the function $\boldsymbol{\gamma}^{k-1}(v,\boldsymbol{\tau})$ are the fixed point linearization where $\boldsymbol{\gamma}^{k-1}(v,\boldsymbol{\tau}):=\underline{\mathbf{B}}(\bar u_h^{k-1},\boldsymbol{\sigma}_h^{k-1})\boldsymbol{\tau}$, and the Newton linearization where

$$\begin{aligned}\boldsymbol{\gamma}^{k-1}(v,\boldsymbol{\tau}):=\underline{\mathbf{B}}(\bar u_h^{k-1},\boldsymbol{\sigma}_h^{k-1})\boldsymbol{\tau}+(v-\bar u_h^{k-1})\partial_v\underline{\mathbf{B}}(\bar u_h^{k-1},\boldsymbol{\sigma}_h^{k-1})\boldsymbol{\sigma}_h^{k-1}\\+(\partial_{\boldsymbol{\tau}}\underline{\mathbf{B}}(\bar u_h^{k-1},\boldsymbol{\sigma}_h^{k-1})\cdot\boldsymbol{\sigma}_h^{k-1})\cdot(\boldsymbol{\tau}-\boldsymbol{\sigma}_h^{k-1}).\end{aligned} \tag{7.13}$$

Problem (7.12) gives rise to a linear system which is of a saddle-point form, written for a couple of vectors associated with the discrete functions $\bar{u}_h^k$ and $\boldsymbol{\sigma}_h^k$. As such, it is not suitable to the present framework. However, following [37, 40], the resulting algebraic systems can be equivalently rewritten in the form (7.2). In particular, any appearance of the flux unknown $\boldsymbol{\sigma}_h^k$ is eliminated, and the only unknowns are the discrete potentials $\bar{u}_h^k$. Then, the approach of Section 7.3-Section 7.4 can be readily used.

8 Conclusions

In this work, we have designed an inexact Newton method with adaptive stopping criteria for the iterative nonlinear and linear solvers. These criteria are based on guaranteed and robust a posteriori error estimates. A complete adaptive strategy combined with adaptive mesh refinement has also been proposed. We have presented numerical experiments illustrating the computational gains achieved by our approach. Our error estimates are derived in an abstract unified framework using equilibrated flux reconstructions. These reconstructions must comply with two assumptions which we have verified for a wide class of discretization schemes and linearizations. In some cases, local mixed finite element problems are to be solved. In practice, the corresponding local matrices can be assembled only once in a preprocessing stage. Additional computational savings are possible by evaluating the error estimators only periodically and not at each iteration of both solvers. Flux reconstructions on more complex meshes (possessing hanging nodes and elements of general shapes) are the subject of ongoing work.

References

[1] M. Ainsworth, *A framework for obtaining guaranteed error bounds for finite element approximations*, J. Comput. Appl. Math., 234 (2010), pp. 2618–2632.

[2] T. Arbogast and Z. Chen, *On the implementation of mixed methods as nonconforming methods for second-order elliptic problems*, Math. Comp., 64 (1995), pp. 943–972.

[3] M. Arioli, D. Loghin, and A. J. Wathen, *Stopping criteria for iterations in finite element methods*, Numer. Math., 99 (2005), pp. 381–410.

[4] R. Becker, C. Johnson, and R. Rannacher, *Adaptive error control for multigrid finite element methods*, Computing, 55 (1995), pp. 271–288.

[5] D. Braess and J. Schöberl, *Equilibrated residual error estimator for edge elements*, Math. Comp., 77 (2008), pp. 651–672.

[6] F. Brezzi and M. Fortin, *Mixed and hybrid finite element methods*, vol. 15 of Springer Series in Computational Mathematics, Springer-Verlag, New York, 1991.

[7] A. Buffa and C. Ortner, *Compact embeddings of broken Sobolev spaces and applications*, IMA J. Numer. Anal., 29 (2009), pp. 827–855.

[8] E. Burman and A. Ern, *Discontinuous Galerkin approximation with discrete variational principle for the nonlinear Laplacian*, C. R. Math. Acad. Sci. Paris, 346 (2008), pp. 1013–1016.

[9] C. Carstensen, *A unifying theory of a posteriori finite element error control*, Numer. Math., 100 (2005), pp. 617–637.

[10] C. Carstensen and R. Klose, *A posteriori finite element error control for the p-Laplace problem*, SIAM J. Sci. Comput., 25 (2003), pp. 792–814.

[11] A. L. Chaillou and M. Suri, *Computable error estimators for the approximation of nonlinear problems by linearized models*, Comput. Methods Appl. Mech. Engrg., 196 (2006), pp. 210–224.

[12] ——, *A posteriori estimation of the linearization error for strongly monotone nonlinear operators*, J. Comput. Appl. Math., 205 (2007), pp. 72–87.

[13] E. Creusé, M. Farhloul, and L. Paquet, *A posteriori error estimation for the dual mixed finite element method for the p-Laplacian in a polygonal domain*, Comput. Methods Appl. Mech. Engrg., 196 (2007), pp. 2570–2582.

[14] P. Destuynder and B. Métivet, *Explicit error bounds for a nonconforming finite element method*, SIAM J. Numer. Anal., 35 (1998), pp. 2099–2115.

[15] ———, *Explicit error bounds in a conforming finite element method*, Math. Comp., 68 (1999), pp. 1379–1396.

[16] D. A. Di Pietro and A. Ern, *Discrete functional analysis tools for discontinuous Galerkin methods with application to the incompressible Navier-Stokes equations*, Math. Comp., 79 (2010), pp. 1303–1330.

[17] D. A. Di Pietro and A. Ern, *Mathematical Aspects of Discontinuous Galerkin Methods*, vol. 69 of Mathématiques & Applications, Springer-Verlag, Berlin, 2011.

[18] L. El Alaoui, A. Ern, and M. Vohralík, *Guaranteed and robust a posteriori error estimates and balancing discretization and linearization errors for monotone nonlinear problems*, Comput. Methods Appl. Mech. Engrg., 200 (2011), pp. 2782–2795.

[19] A. Ern, S. Nicaise, and M. Vohralík, *An accurate* $\mathbf{H}(\mathrm{div})$ *flux reconstruction for discontinuous Galerkin approximations of elliptic problems*, C. R. Math. Acad. Sci. Paris, 345 (2007), pp. 709–712.

[20] A. Ern and M. Vohralík, *Flux reconstruction and a posteriori error estimation for discontinuous Galerkin methods on general nonmatching grids*, C. R. Math. Acad. Sci. Paris, 347 (2009), pp. 441–444.

[21] ———, *A posteriori error estimation based on potential and flux reconstruction for the heat equation*, SIAM J. Numer. Anal., 48 (2010), pp. 198–223.

[22] R. Eymard, T. Gallouët, and R. Herbin, *Finite volume methods*, in Handbook of Numerical Analysis, Vol. VII, North-Holland, Amsterdam, 2000, pp. 713–1020.

[23] ———, *Finite volume approximation of elliptic problems and convergence of an approximate gradient*, Appl. Numer. Math., 37 (2001), pp. 31–53.

[24] W. Han, *A posteriori error analysis via duality theory*, vol. 8 of Advances in Mechanics and Mathematics, Springer-Verlag, New York, 2005. With applications in modeling and numerical approximations.

[25] A. Hannukainen, R. Stenberg, and M. Vohralík, *A unified framework for a posteriori error estimation for the Stokes problem*, Numer. Math., (2012). Accepted for publication.

[26] P. Houston, E. Süli, and T. P. Wihler, *A posteriori error analysis of hp-version discontinuous Galerkin finite-element methods for second-order quasi-linear elliptic PDEs*, IMA J. Numer. Anal., 28 (2008), pp. 245–273.

[27] P. Jiránek, Z. Strakoš, and M. Vohralík, *A posteriori error estimates including algebraic error and stopping criteria for iterative solvers*, SIAM J. Sci. Comput., 32 (2010), pp. 1567–1590.

[28] K. Y. Kim, *A posteriori error estimators for locally conservative methods of nonlinear elliptic problems*, Appl. Numer. Math., 57 (2007), pp. 1065–1080.

[29] A. H. Le, *A posteriori error estimation for simulation of diffusion and fluid mechanics problems by finite volume techniques*, Ph.D. thesis, Université Paris 13, 2011.

[30] R. Luce and B. I. Wohlmuth, *A local a posteriori error estimator based on equilibrated fluxes*, SIAM J. Numer. Anal., 42 (2004), pp. 1394–1414.

[31] L. D. Marini, *An inexpensive method for the evaluation of the solution of the lowest order Raviart–Thomas mixed method*, SIAM J. Numer. Anal., 22 (1985), pp. 493–496.

[32] F. A. Milner, *Mixed finite element methods for quasilinear second-order elliptic problems*, Math. Comp., 44 (1985), pp. 303–320.

[33] J. Pousin and J. Rappaz, *Consistency, stability, a priori and a posteriori errors for Petrov-Galerkin methods applied to nonlinear problems*, Numer. Math., 69 (1994), pp. 213–231.

[34] W. Prager and J. L. Synge, *Approximations in elasticity based on the concept of function space*, Quart. Appl. Math., 5 (1947), pp. 241–269.

[35] S. I. Repin, *A posteriori error estimation for nonlinear variational problems by duality theory*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI), 243 (1997), pp. 201–214, 342.

[36] R. Verfürth, *A posteriori error estimates for nonlinear problems. Finite element discretizations of elliptic equations*, Math. Comp., 62 (1994), pp. 445–475.

[37] M. Vohralík, *Equivalence between lowest-order mixed finite element and multi-point finite volume methods on simplicial meshes*, M2AN Math. Model. Numer. Anal., 40 (2006), pp. 367–391.

[38] ——, *Residual flux-based a posteriori error estimates for finite volume and related locally conservative methods*, Numer. Math., 111 (2008), pp. 121–158.

[39] ——, *Unified primal formulation-based a priori and a posteriori error analysis of mixed finite element methods*, Math. Comp., 79 (2010), pp. 2001–2032.

[40] M. Vohralík and B. I. Wohlmuth, *Mixed finite element methods: implementation with one unknown per element, local flux expressions, positivity, polygonal meshes, and relations to other methods.* Preprint R10031, Laboratoire Jacques-Louis Lions and HAL Preprint 00497394, submitted for publication, 2010.